Characterization of microbial community members and catabolic genes selected for performance under pollutant stress

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades einer
Doktorin der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Daiana de Lima Morales
aus Porto Alegre, Brazil

1. Referent: apl. Professor Dr. Dietmar H. Pieper

2. Referentin: Professorin Dr. Susanne Engelmann

eingereicht am: 06.08.2014

mündliche Prüfung (Disputation) am: 07.11.2014

Druckjahr 2014

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

de Lima-Morales D, Chaves-Moreno D, Jarek M, Vilchez-Vargas R, Jauregui R & Pieper DH (2013) Draft genome sequence of *Pseudomonas veronii* strain 1YdBTEX2. *Genome Announc* **1**: e00258-13.

Tagungsbeiträge

Morales D, Vilchez-Vargas R & Pieper DH (2010) Catabolic gene landscape of novel benzene degrading Pseudomonas strains. (Oral Presentation) DGHM/VAAM– Hannover – Deutschland

Morales D, Chaves D, Vilchez-Vargas R, Jáuregui R & Pieper DH (2011) The genome of two *Pseudomonas veronii* Strains Crucial for Benzene degradation in a Benzene Contaminated Site. (Poster) ProkaGENOMICS – Göttingen – Deutschland

Morales D, Chaves D, Vilchez-Vargas R, Jáuregui R & Pieper DH (2012) The genome of novel benzene degrading *Pseudomonas veronii* strains. (Oral Presentation) Pollutant biodegradation under environmental stress: towards rational bioaugmentation – Amsterdam – The Netherlands

Acknowledgements

I would like to thank:

My supervisor Prof. Dr. Dietmar Pieper, for giving me the opportunity to be part of a great project, for sharing his impressive knowledge and for all his help and commitment, not only with the work but also with me as his student.

Ramiro for his supervision. The MINP group members, you all enriched my life in some way. Silke and Iris for the help in the lab and representing the german culture. Melissa for keeping me motivated, Agnes for her sweet company, Diego for his professional wisdom and irreplaceable friendship, Amelia for her lessons, friendship and admirable strength.

Friends, specially Andreia, Marcelo, Ana, Bahram, Marcela, Marcelinho, Alex, Sagra, Monia, Andrew and my second family Rolf and Felicia thank you for making my stay in Germany easier.

To my sister, I couldn't pass through this without her... Not only the PhD, but my entire life. Meu companheiro Gabi, sempre me surpreendendo e divertindo, nunca canso da sua companhia. Pai e mãe sempre comigo, perto ou longe, obrigada por sustentarem os meus pilares.

Finally thanks to Germany for the life experience…

"There is a special place in my heart for the ones who were with me at my lowest and still loved me when I wasn't very loveable."

Table of Contents

Chapter I - General Introduction.....9
1.1 BTEX pollution – the compounds and the problem.....9
1.2.1 Bacterial metabolism of BTEX – the Upper pathway.....14
1.2.2 Bacterial metabolism of BTEX – the Lower pathway.....17
1.2.3 Enzyme families crucial for aromatic degradation.....18
1.3 Investigating isolates, communities and catabolic genes involved in biodegradation.....29
1.3.1 Current molecular fingerprinting tools at our disposal.....29
1.4 Summary of techniques employed in this work.....36
1.5 Research approach.....38
1.6 Specific aims.....38
1.7 References.....39

Chapter II – Genome Sequence of the soil bacterium *Pseudomonas veronii* strain 1YdBTEX2......52
2.1 Abstract.....53
2.2 Genome Announcement.....53
2.3 Nucleotide sequence accession numbers.....54
2.4 Acknowledgements.....55
2.5 References.....55

Chapter III - Physiological diversity of *Pseudomonas veronii* 1YdBTEX2 and 1YB2, two versatile benzene degrading microorganisms.....57
3.1 Abstract 58
3.2 Introduction.....59
3.3 Material and Methods.....60
3.3.1 Strains and culture conditions.....60
3.3.2 Genome sequencing and assembling.....61
3.3.3 Phylogenetic comparisons.....61
3.3.4 Partial purification of benzene dihydrodiol dehydrogenase.....62
3.3.5 Identification of benzene dihydrodiol dehydrogenase.....63
3.3.6.Preparation of standards for real-time quantitative PCR (RT-q-PCR).....63
3.3.7 RNA extraction and reverse transcription.....65
3.3.8 Real Time Quantitative PCR (RT-q-PCR).....65
3.3.9 Acession Number.....65
3.4 Results and Discussion.....65

3.4.1 General characteristics of *P. veronii* 1YdBTEX2 and 1YB2.....65
3.4.2 Phylogenetic analyses.....66
3.4.3 The catabolic potential of *P. veronii* 1YdBTEX2 and 1YB2.....74
3.4.4 Genetic attributes that may help *Pseudomonas veronii* 1YdBTEX2 and 1YB2 to survive under environmental conditions.....82
3.5 Conclusions.....89
3.6 Acknowledgements.....89
3.7 References.....90

Chapter IV – Microbial community and catabolic gene structure in soil under benzene and BTEX stress.....97
4.1 Abstract.....99
4.2 Introduction.....98
4.3 Materials and Methods.....101
4.3.1.Samples.....101
4.3.2 Experiment design.....102
4.3.3 Catabolic gene analysis.....103
4.3.4 Community analysis.....105
4.3.5 Bioinformatic analysis.....107
4.4 Results and Discussion.....109
4.4.1 Soil community structure.....109
4.5 Conclusions.....120
4.6 Acknowledgements.....120
4.7 References.....121

Chapter V - General Discussion.....127
5.1 General Discussion.....127
5.2 References.....136

6 Appendix.....140

Summary

Benzene and the structurally related toluene, ethylbenzene and xylenes, collectively known as BTEX, are important environmental contaminants. In the current thesis I aimed to sequence and assemble the genomes of two key players for benzene degradation at a contaminated site in order to elucidate the organization of the pathways and reveal why these isolates prevail in the environment under pollutant stress. More widely, I analyzed microbial community and catabolic gene structures of contaminated soils from different origins under constant benzene and BTEX contamination.

The genomes of *P. veronii* 1YdBTEX2 and *P. veronii* 1YB2 were sequenced, assembled and annotated. Comparison of the genomes with those of previously sequenced strains indicated that they belong to a species different from *Pseudomonas fluorescens*, and analysis of phylogenetic marker genes showed that they are in fact *Pseudomonas veronii* isolates. Whole genome comparisons then gave indications into the phylogeny of species of the *Pseudomonas fluorescens* group.

Analysis of the benzene catabolic pathways revealed the plasticity of those genomes in order to perform efficient degradation. It revealed that they harbor unique catabolic pathways comprising a gene cluster encoding an isopropylbenzene dioxygenase where genes encoding downstream enzymes were interrupted by stop codons. Extradiol dioxygenases were recruited from gene clusters comprising genes encoding a 2-hydroxymuconic semialdehyde dehydrogenase necessary for benzene degradation but typically absent from isopropylbenzene dioxygenase encoding gene clusters. The benzene dihydrodiol dehydrogenase encoding gene was not clustered with any other aromatic degradation gene and the encoded protein was only distantly related to dehydrogenases of aromatic degradation pathways. The involvement of the different gene clusters in the degradation pathways was verified through RT-PCR. Both *Pseudomonas veronii* strains harbour unique features such as gene clusters for the synthesis of the siderophore pyochelin, a complete denitrification pathway organized in one superoperonic cluster and a gene cluster comprising a [NiFe]-hydrogenase for the use of hydrogen as alternative electron donor. These genetic attributes may help *P. veronii* 1YdBTEX2 and 1YB2 to survive under environmental conditions.

In order to identify commonalities or differences in the adaptation of microbial communities to pollutant stress, three types of soil were subjected to a constant supply of benzene or BTEX and surveyed by high throughput methods for microbial community and catabolic gene structures. Whereas soil previously subjected to intensive *in-situ* bioremediation showed only negligible changes in community structure, other contaminated soil samples showed a clear succession of phylotypes. A rapid response to benzene stress was observed whereas the response to BTEX pollution was significantly slower. After extended incubation, actinobacterial phyloypes were increasing in relative abundance, indicating their superior fitness to pollution stress. The observed shifts in the microbial communities could be correlated with shifts in the composition of catabolic genes, which might be involved in the catabolism of benzene/BTEX and selected under respective pollutant stress.

Overall, this work contributes to a better understanding on bacteria becoming dominant under pollutant stress and expands the knowledge on bacterial community performance under unfavorable conditions.

Zusammenfassung

Benzol, Toluol, Ethylbenzol und Xylole (BTEX) sind wichtige Umweltschadstoffe. In der vorliegenden Arbeit sollten die Genome von zwei Schlüsselorganismen des Benzolabbaus in einer kontaminierten Umwelt sequenziert und assembliert werden, um die Organisation des Abbauweges aufzuklären und der Frage nachzugehen, aus aus welchem Grunde diese in der Umwelt unter Schadstoffstress einen Vorteil haben. Darüber hinaus analysierte ich die Zusammensetzung der mikrobiellen Gemeinschaft und der Gen-Strukturen des Schadstoffabbaus kontaminierter Böden bei gleichbleibender Belastung mit Benzol und BTEX.

Die Genome von *Pseudomonas veronii* 1YdBTEX2 und 1YB2 wurden sequenziert, assembliert und annotiert. Vergleich der Genome mit denen zuvor sequenzierter Stämme deutete darauf hin, dass sie zu einer von *P. fluorescens* verschiedenen Art gehören und die phylogenetische Analyse zeigte, dass sie in der Tat zur Art *P. veronii* gehören. Genom-Vergleiche gaben darüber hinaus Hinweise auf die Phylogenie der Arten der *P. fluorescens* Gruppe.

Die Analyse der Abbauwege für Benzol zeigte die Plastizität der Genome. Die Stämme haben einzigartige Abbauwege, die ein für eine Isopropylbenzol Dioxygenase kodierendes Gen-Cluster, in dem die für nachgeschaltete Enzyme kodierenden Gene durch Stop-Codons unterbrochen werden, beinhalten. Extradiol Dioxygenasen wurden aus Gen-Clustern rekrutiert, die über ein für eine 2-Hydroxymuconsäuresemialdehyd Dehydrogenase kodierendes Gen verfügen, welches für den Benzolabbau unabdingbar ist. Die unabhängig von diesen Genen kodierte Benzoldihydrodiol Dehydrogenase wies nur geringe Ähnlichkeit mit bisher beschriebenen Enzymen des Aromatenabbaus auf. Die Wichtigkeit der verschiedenen Gene für den Abbau wurde durch RT-PCR verifiziert. Beide Stämme beherbergen zudem für ungewöhnliche Eigenschaften kodierende Gene wie ein Gen-Cluster zur Synthese des Siderophors Pyochelin, Gene für Enzyme zur vollständigen Denitrifikation, die in einem superoperonischen Cluster organisiert sind und ein Gen-Cluster welches für eine [NiFe]-Hydrogenase zur Nutzung von Wasserstoff als alternativem Elektronenspender kodierende Gene beinhaltet. Diese genetischen Attribute können *P. veronii* 1YdBTEX2 und 1YB2 helfen, unter widrigen Umweltbedingungen zu überleben.

Um Gemeinsamkeiten oder Unterschiede in der Anpassung mikrobieller Gemeinschaften an Schadstoffstress zu analysieren, wurden drei Bodentypen konstant mit Benzol/BTEX inkubiert und mittels Hochdurchsatzmethoden die Zusammensetzung der mikrobiellen Gemeinschaft wie auch der katabolischen Gene untersucht. Während der einer in-situ Sanierung unterworfene Boden nur geringfügige Veränderungen in der Gemeinschaftsstruktur zeigte, wiesen andere kontaminierte Bodenproben eine klare Abfolge von Phylotypen auf. Nach längerer Inkubation nahmen Actinobacterium-Phylotypen deutlich in ihrer relativen Häufigkeit zu, was auf ihre überlegene Fitness gegenüber Schadstoffstress hindeutet. Die beobachteten Veränderungen in den mikrobiellen Gemeinschaften konnten mit Verschiebungen in der Zusammensetzung von in den Abbau von BTEX/Benzol involvierten Genen korreliert werden.

Insgesamt trägt diese Arbeit zu einem besseren Verständnis von Mikroorganismen, die sich unter Schadstoffstress durchsetzen, bei und erweitert das Wissen über die Gemeinschaftsleistungen von Bakterien unter ungünstigen Umweltbedingungen.

Chapter I - General Introduction

1.1 BTEX pollution – the compounds and the problem

Volatile organic compounds and hazardous air pollutants are important environmental contaminants. Among these are benzene and the structurally related toluene, ethylbenzene and the isomeric o-, m- and p-xylenes, collectively known as BTEX (Figure 1). They are the major aromatic hydrocarbons in petroleum products like gasoline, diesel fuel, lubricating and heating oil. BTEX are volatile, usually colorless with strong odors and highly flammable. They are extensively used in industry as solvents for organic synthesis and the manufacture of chemicals, rubber, plastics, in solvent paints and lacquers, for equipment cleaning and for other downstream processing purposes. BTEX use is widespread and prevalent.

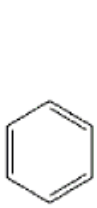
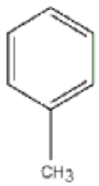
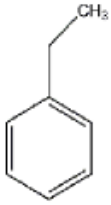
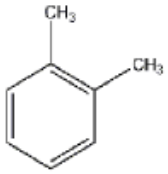
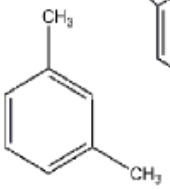
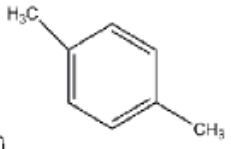
Name	Benzene	Toluene	Ethylbenzene	o-Xylene	m-Xylene	p-Xylene
Chemical structure						
Molecular formula	C_6H_6	C_7H_8	C_8H_{10}	C_8H_{10}		
Water solubility	1780 mg l^{-1}	535 mg l^{-1}	152 mg l^{-1}	175 mg l^{-1}	135 mg l^{-1}	198 mg l^{-1}
Vapor pressure	76 mmHg	22 mmHg	7 mmHg	5 mmHg	6 mmHg	6.5 mmHg

Figure 1: General characteristics of BTEX

The main sources of BTEX are the chemical industries, fuel companies and vehicular traffic (Fujita et al., 2011; Majumdar, 2011 Correa et al., 2012). They are also frequently discharged into soil and groundwater as a result of leakage through underground storage tanks and pipelines, improper waste disposal practices, inadvertent spills and leaching from landfills (Corseuil et al., 1997; Shim & Yang, 1999; Mao et al., 2008), (http://www.gsi-net.com/files/papers/meamodel.pdf) posing massive environmental threats. For example, in Romania, an accident at an oil refinery in 2002 spread contaminants over 5 km around the

factory (http://whqlibdoc.who.int/publications/2006/9241546689_eng.pdf). Also, many severe oil spills in the ocean are well documented. The International Tanker Owners Pollution Federation Limited (http://www.itopf.com/information-services/data-and-statistics/ statistics/) reports that from 1970 until 2012, nearly 10,000 incidents occurred and approximately 5.75 million tons of oil was lost as a result of tanker incidents. This number doesn't take into account every kind of accident involving petrol, but only those related to oil barrels transported by ship, such that the actual pollution is even higher. *The Deepwater Horizon Oil Spill* was a recent catastrophic petrol spill in the Gulf of Mexico (in 2010), releasing around 4.9 million barrels of crude oil into the deep ocean for more than 200 days until the well was closed (http://www.restorethegulf.gov/sites/default/files/documents/pdf/OilBudgetCalc_Full_HQ-Print_111110.pdf).

As other volatile organic compounds, BTEX compounds are involved in the formation of ground level ozone which can damage crops and materials (Bravo-Linares & Mudge, 2007). Since BTEX are volatile, most releases end up in the atmosphere, although some can be bound to soils and sediments (Bravo-Linares & Mudge, 2007).

In Europe, BTEX are responsible for 6% of the total contamination of soils (Figure 2). The World Health Organization (WHO) reports BTEX together with chlorinated hydrocarbons to be the most relevant groundwater contaminants, with significant impact on drinking water (Figure 3). BTEX are among the most hydrophilic petroleum compounds (see water solubility in Figure 1), such that they move relatively easily through the water column, rapidly reaching groundwater reservoirs (Bertin et al., 2008). So, while BTEX are naturally occurring compounds in the environment at negligible levels, our insatiable desire and need to use products that contain BTEX presents increased risks for environmental spills and contamination.

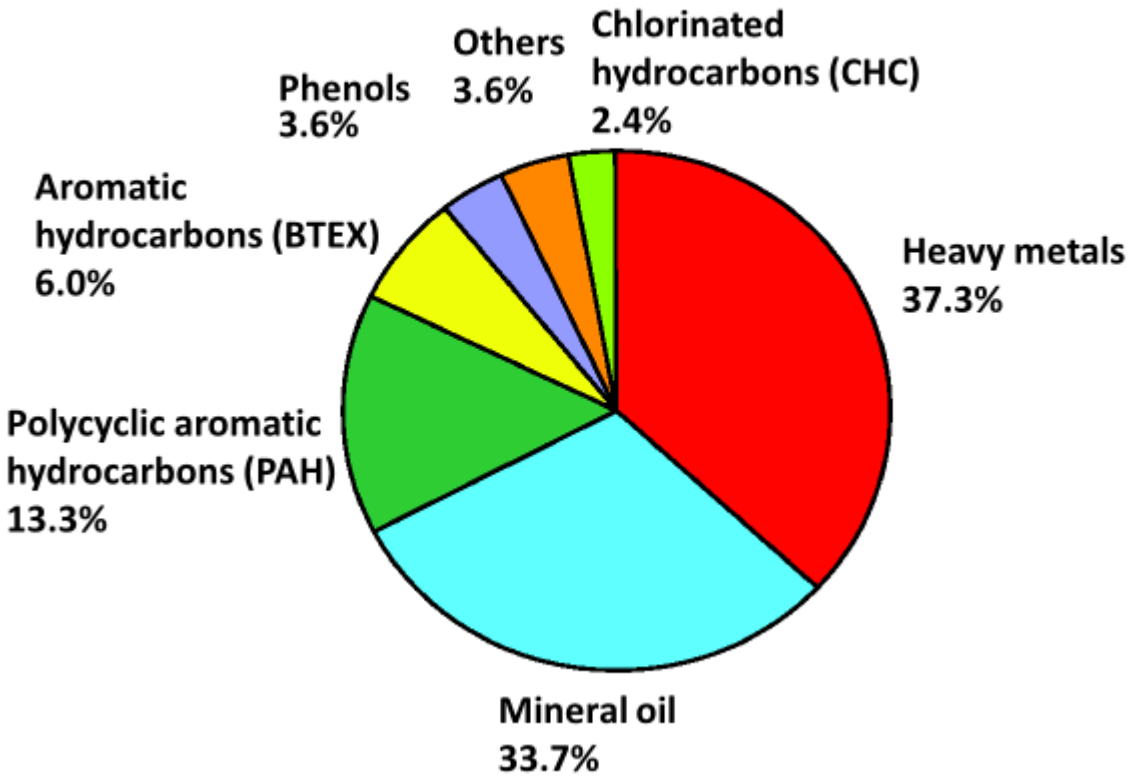


Figure 2: Overview of contaminants affecting soil in Europe, according to the European Environment Agency (http://www.eea.europa.eu).

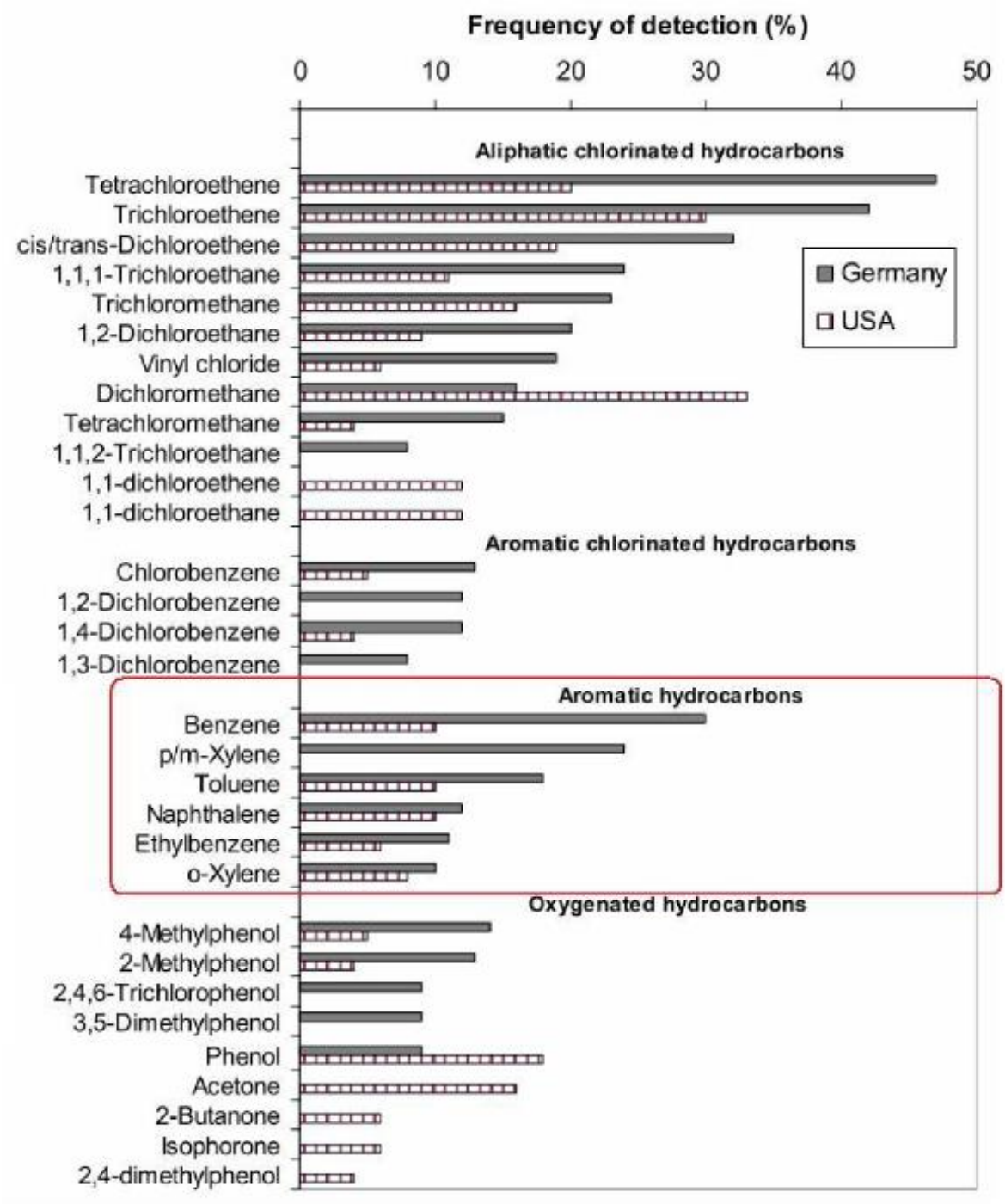


Figure 3: The 25 most frequently detected organic groundwater contaminants at hazardous waste sites in Germany (250 sites) and the USA (500 sites) (based on concentrations ≥ 1µg/l). BTEX are highlighted in red (Protecting Groundwater for Health, WHO document, adapted from Kerndorff et al., 1992).

Regarding benzene, many countries already have regulations for its levels in the atmosphere. The European Commission adopted an air quality objective of 5 μgm^{-3} (annual

average). In the US, the National Institute for Occupational Safety and Health (NIOSH) established occupational exposure limits to BTEX in the air. Exposure to benzene is very restrictive, and should not be higher than 0.319 mg m^{-3} on average during 8 hours per day, whereas the limits for toluene and ethylbenzene/xylenes are 375 mg m^{-3} and 435 mg m^{-3} respectively.

Long-term exposure to high concentrations of BTEX, which is usually only experienced in occupational settings leads to damage of the liver, kidneys, central nervous system and eyes. More specifically, benzene is classified as a "Group A, known human carcinogen of medium carcinogenic hazard" by the US EPA (Pollutants et al., 1995). Benzene is also now considered teratogenic, and associated with the development of leukemia (Schnatter et al., 2005). However, the other BTEX compounds, toluene, xylenes and ethylbenzene (TEX) are classified as "Group D, not classifiable as to human carcinogenicity" by the US EPA (Pollutants et al., 1995). Nevertheless exposure to TEX is harmful and inhalation can cause irritation of the upper respiratory tract and eyes, sore throat and dizziness. Acute exposure to toluene and xylenes by inhalation or absorption through skin can cause central nervous system dysfunction and narcosis (Shim et al., 2002). While the main risk for humans is through inhalation, the other concern is drinking water where pollution can be substantial. In fact there are guidelines regulating BTEX levels in the environment, however, the control of BTEX in the environment is still not obligatory in every country.

Concerning drinking-water the US EPA has established permissible levels of BTEX in the water supply systems. The benzene permissible level is 0.005 mg/liter, whereas those for the less toxic toluene, ethylbenzene and total xylenes are 1, 0.7 and 10 mg/liter respectively. In addition the release of BTEX compounds is controlled through the OSPAR convention which protects the marine environment of the north-east Atlantic sea (http://www.ospar.org/), the UNECE convention on long-range transboundary air pollution (http://www.unece.org/), the Basel convention on transboundary movements and disposal of hazardous wastes (http://www.basel.int/convention/basics.html) and the National Air Quality Strategy regulations and pollution of surface waters in the UK (http://www.environment-agency.gov.uk) among others.

1.2 BTEX pollution – a solution through remediation

Among all remediation technologies for treating contaminated sites, biological processes are economical, energy-efficient and environmentally sound approaches (Lapertot et al., 2007). During previous years, natural attenuation, defined specifically as "naturally occurring processes in soil and groundwater that act without human intervention to reduce the mass, toxicity, mobility, volume or concentration of contaminants in those media" has received increasing attention. It is generally accepted that microorganisms are the principal mediators of the natural attenuation of many pollutants. However, in some cases pollutants may be transformed into more toxic products, as reported for the anaerobic transformation of trichloroethylene (Bagley & Gossett, 1990). The use of natural attenuation thus requires detailed monitoring to determine how effective this method is for attaining site remediation goals. As is the case for natural attenuation, an understanding of the response of the indigenous microbial community is also necessary for successful stimulation. Even more unpredictable are bioaugmentation approaches, as the added microorganisms, which are supposed to increase the degradative performance on site, not only have to express these capabilities under *in situ* conditions, but also have to compete with the complex microbiota inhabiting the site under study. It is thus evident that only a detailed understanding of the functioning and interactions within microbial communities will allow their rational manipulation and overcome factors that limit efficient bioremediation.

1.2.1 Bacterial metabolism of BTEX – the Upper pathway

Some microorganisms are able to use aromatic compounds such as BTEX as their carbon source. The group of enzymatic reactions involved in BTEX degradation can be divided into the upper pathway, where the key step is the ring activation and the lower pathway, where the ring cleavage is the key step.

The first step in the degradation of BTEX is to overcome the chemical stability of the aromatic ring. This step is referred to as ring activation. Three different pathways can be employed by bacteria to activate aromatic rings: 1) Activation by a dioxygenase followed by a dehydrogenase reaction to form the intermediate catechol structure 2) activation by two

successive monooxygenase reactions with a phenolic intermediate and 3) activation by monooxygenation of the methyl function to give benzylalcohol (Figure 4).

Benzene, toluene and ethylbenzene can be activated by a dioxygenase resulting in the formation of benzene-cis-dihydrodiol, toluene-*cis*-dihydrodiol) and ethylbenzene-*cis*-dihydrodiol, respectively (Gibson & Parales, 2000). These structures are then dehydrogenated to catechol, 3-methylcatechol and 3-ethylcatechol, respectively. The dioxygenases that initiate the degradation of aromatics are called Rieske non-heme iron oxygenases (ring hydroxylating dioxygenases, RHDO) (Gibson & Parales, 2000).

Benzene, toluene and o-xylene can be activated by two successive monooxygenations. The monooxygenation of the benzene ring by a benzene monooxygenase results in a phenol structure, which is further transformed to a catechol by a phenol monooxygenase (also termed phenol hydroxylase). Monooxygenation of toluene results in the formation of o-cresol (Shields et al., 1989), p-cresol (Whited & Gibson, 1991) (Figure 4) or m-cresol (Olsen et al., 1994). However, later analysis has shown that the toluene monooxygenase of *R. pickettii* PK01, which had been reported previously to hydroxylate toluene at the meta position, producing primarily 3-methylphenol (Olsen et al., 1994), hydroxylates toluene predominantly at the para position producing 4-methylphenol (Fishman et al., 2004). These intermediates are further transformed to 3-methylcatechol (from o-cresol) and protocatechuate (from p-cresol), protocatechuate is then oxidized by ortho ring cleavage (see figure 4).

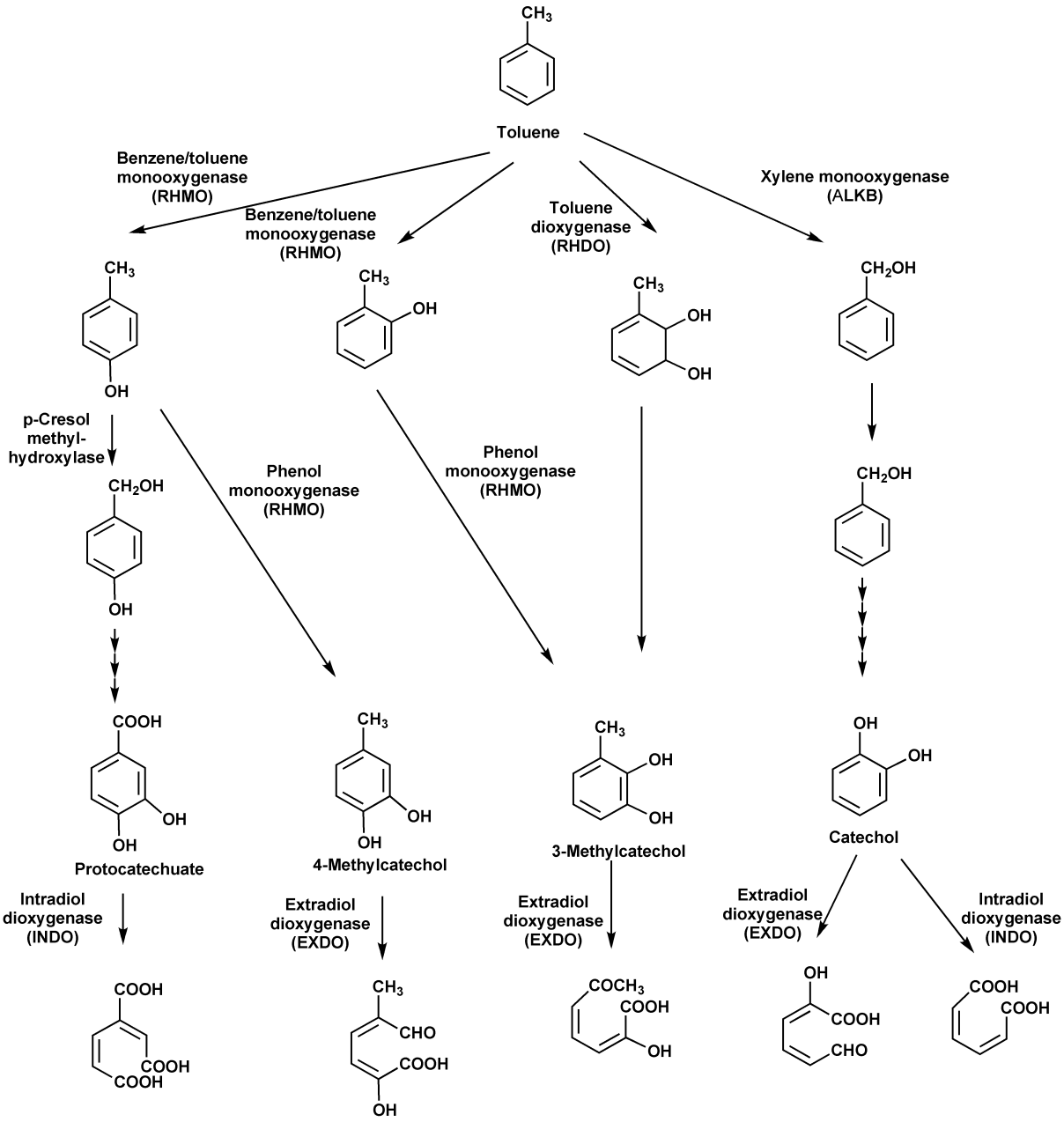


Figure 4: Pathways for toluene degradation. Enzyme superfamilies that initiate the degradation are soluble di-iron monooxygenases (RHMO), Rieske non-heme iron oxygenases (RHDO) and integral membrane-bound hydroxylases (ALKB). Enzyme superfamilies catalyzing ring-cleavage are extradiol dioxygenases of the vicinal chelate superfamily (EXDO) and intradiol dioxygenases (INDO).

The monooxygenases that initiate the degradation of monoaromatics belong to the soluble diiron monooxygenases (RHMO) (Leahy et al., 2003). Alternative to oxygenation of the aromatic ring, degradation can be initiated by monooxygenation of the methyl substituent of toluene, m-xylene and p-xylene. Monooxygenation of the methyl function

results in the formation of (methyl)benzylalcohols (Worsey & Williams, 1975) which may be further dehydrogenated and oxidized to benzoate, or 3- or 4-methylbenzoate, respectively. By dioxygenation and dehydrogenation, these intermediates are converted to catechol (from toluene, see Figure 4) 4-methylcatechol (from p-xylene) or 3-methylcatechol (from m-xylene).

1.2.2 Bacterial metabolism of BTEX – the lower pathway

Subsequent to the ring activation the next key step in BTEX degradation is the ring cleavage. There are two ways to cleave the central catechol, protocatechuate and catechol derivatives by 1) an intradiol dioxygenase, breaking the ring between adjacent carbons or by 2) an extradiol dioxygenase, breaking the ring between a hydroxylated and an adjacent carbon atom. The respective ring-cleavage products are further processed to Krebs cycle intermediates (Figure 4).

Intradiol dioxygenases: The catechol 1,2-dioxygenase and the protocatechuate 3,4-dioxygenase belong to the Intradiol dioxygenase enzyme superfamily. They transform catechol to muconate and protocatechuate to 3-carboxymuconate, respectively (Pérez-Pantoja et al., 2009). These enzymes harbor non-heme Fe (III) as a prosthetic group (Hayaishi, 1950; Hayaishi et al., 1957). Catechol 1,2-dioxygenases have a relaxed substrate specificity, thus being capable of transforming methyl substituted derivatives (Dorn & Knackmusst, 1978; Murakami et al., 1997), however, mineralization of methylcatechols is only seldomly achieved via this route (Pieper et al., 1985).

Extradiol dioxygenases: Usually the degradation of BTEX results in the formation of catechol or substituted catechols as the central intermediates that are degraded by an extradiol dioxygenase (Figure 4). The majority of extradiol dioxygenases belong to the vicinal chelate superfamily and contain a non-heme Fe(II) in the active site. This enzyme superfamily comprises among others catechol 2,3-dioxygenases, that catalyze the transformation of catechol to 2-hydroxymuconic semialdehyde, actinobacterial homoprotocatechuate 2,3-

dioxygenases, that catalyze the transformation of homoprotocatechuate to 5-carboxymethyl-2-hydroxymuconic semialdehyde, 2,3-dihydroxybiphenyl 1,2-dioxygenases, that catalyze the transformation of 2,3-dihydroxybiphenyl to 2-hydroxy-6-phenyl-6-oxohexa-2,4-dienoate and 1,2-dihydroxynaphthalene dioxygenases that transform naphthalene-1,2-diol to 2-hydroxychromene-2-carboxylate. They can be divided into two large groups according to their substrate preference, either preferring monocyclic or bicyclic aromatic structures (Eltis & Bolin, 1996).

1.2.3 Enzyme families crucial for aromatic degradation

Genes involved in aromatic degradation have common evolutionary origins. Comparisons of the amino acid sequences of the encoded enzymes reveal that they share consensus sequences and belong to enzyme superfamilies of diverse but evolutionarily related enzymes. The enzyme sequences have been retrieved from the NCBI public database (http://www.ncbi.nlm.nih.gov), classified and validated for the construction of a curated database (Vilchez-Vargas et al., 2013). Those enzyme sequences belonging to 5 superfamilies assumed to be crucial for aromatic degradation were aligned and phylogenetic trees constructed as summarized below.

Rieske Non-heme Iron Oxygenases – RHDO

Rieske non-heme iron oxygenases (RHDO) initiate the degradation of a wide range of environmental pollutants such as benzoate, benzene, toluene, phthalate, naphthalene and biphenyl (Gibson & Parales, 2000) by the addition of two hydroxyl groups to the aromatic nucleus. These enzymes are multicomponent systems typically consisting of a short electron transport chain consisting of a ferredoxin and a ferredoxin reductase, and two subunits α and β. The β-subunit is the structural moiety and the α-subunit is the catalytic site, which defines the substrate specificity of the enzyme (Zielinski et al., 2002). However, some enzymes consist of α-subunits only, and a ferredoxin may be absent from other enzyme systems.

Phylogenetic analyses of the α-subunit sequences from the multicomponent enzymes of the RHDO superfamily indicate that they cluster into families associated with the substrates oxidized by the members forming this superfamily. According to Gibson and Parales (2000) the RHDO superfamily can be divided in at least four major families: 1) the toluene/biphenyl family including enzymes for the degradation of toluene, benzene, isopropylbenzene, chlorobenzenes and biphenyl from both gram-negative and gram-positive organisms; 2) the naphthalene family consisting of enzymes for the degradation of naphthalene and phenanthrene from gram-negative organisms; 3) the benzoate family, a group of enzymes that oxidize aromatic acids and 4) the phthalate family, a large and diverse group of enzymes, that, like enzymes of the benzoate family, oxidize aromatic acids. As information on members of the RHDO superfamily is constantly growing, the phylogeny can be refined and probable substrate specificities can be assumed. In this way Vilchez-Vargas et al. (2013) divided the RHDO superfamily into 21 clusters (Figure 5), where one cluster (cluster-1) comprises enzymes with documented activity as benzene/toluene/isopropylbenzene or biphenyl dioxygenases of proteobacterial as well as actinobacterial origin (Peres-Pantoja et al., 2009). As an example respective enzymes are present in *P. putida* F1 (Gibson et al., 1968) which harbors the *tod* enzymes, in *P. putida* RE204 (Eaton & Timmis, 1986) and *P. fluorescens* IP01 (Eaton, 1996), which harbor the *ipb* enzymes or *B. xenovorans* LB400 (Erickson & Mondello, 1992) and *R. Jostii* RHA1, which harbor the *bph* genes (Masai et al., 1995) (Figure 6-A). The respective degradation pathways typically comprise a multicomponent dioxygenase, a *cis*-dihydrodiol dehydrogenase and an extradiol dioxygenase.

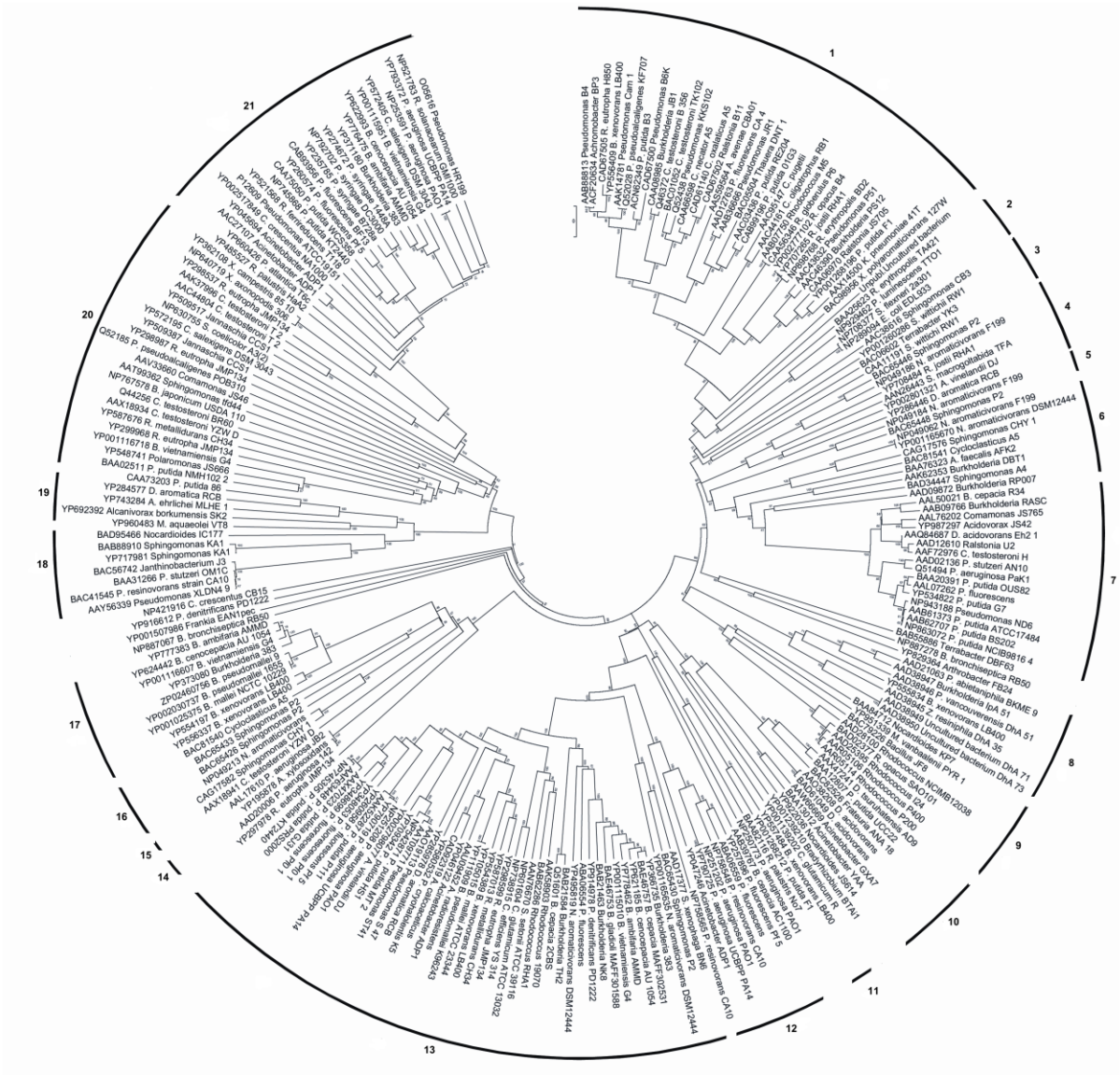


Figure 5: Evolutionary relationships of α-subunits of Rieske non-heme iron oxygenases (RHDO). According to the documented substrate specificity of representative members they can be clustered as follows: Cluster-1, benzene/toluene/isopropylbenzene/biphenyl dioxygenases; cluster-2, 3-phenylpropionate dioxygenases and related enzymes; cluster-3, angular dibenzofuran dioxygenases and related enzymes; cluster-4, tetralin dioxygenases of *S. macrogoltabida* TFA and related enzymes; cluster-5, biphenyl dioxygenases of α-Proteobacteria; cluster-6, phenanthrene dioxygenases from Proteobacteria; cluster-7, proteobacterial naphthalene dioxygenases, cluster-8, diterpenoid dioxygenases; cluster-9, actinobacterial naphthalene/phenanthrene dioxygenases; cluster-10, aniline dioxygenases; cluster-11, p-cumate dioxygenases; cluster-12, anthranilate dioxygenases of γ-Proteobacteria; cluster-13, benzoate dioxygenases; cluster-14, 2-chlorobenzoate dioxygenases; cluster-15, terephthalate dioxygenases; cluster-16, salicylate 1-hydroxylases; cluster-17, anthranilate dioxygenases of *Burkholderia* sp; cluster-18, carbazol dioxygenases; cluster-19, 1-oxo-2,3-dihydroquinone monooxygenases and related enzymes; cluster-20, phthalate 4,5-dioxygenases and cluster- 21, vanillate demethylases.

Typically, the respective operons also comprise a gene encoding a 2-hydroxymuconic semialdehyde hydrolase. Such an enzyme is necessary to process ketones formed after ring-cleavage of 3-methylcatechol, 3-isopropylcatechol or 2,3-dihydroxybiphenyl (Harayama et al., 1987). 2-Hydroxymuconic semialdehyde hydrolase is representing the hydrolytic branch of the meta-cleavage pathway, and both, enzymes of the hydrolytic branch as well as of the 4-oxalocrotonate branch (here 2-hydroxymuconic semialdehyde is oxidized to 2-hydroxymuconate by 2-hydroxymuconic semialdehyde dehydrogenase, followed by isomerization to oxalocrotonate through the action of oxalocrotonate isomerase and decarboxylation by oxalocrotonate decarboxylase to 2-hydroxypent-2,4-dienoate, the common intermediate of the hydrolytic and the 4-oxoalocrotonate branch), are encoded in operons for e.g. salicylate degradation by *Pseudomonas putida* G7 (Yen and Gunsalus 1981) (Figure 6-B), 3-methylbenzoate by *Pseudomonas putida* KT2440 (Harayama et al., 1987) or phenol by *C. necator* JMP134 (Pérez-Pantoja et al., 2008). However, enzymes of the 4-oxalocrotonate branch, which is necessary for the degradation of catechols are typically absent from operons encoding Rieske non-heme iron oxygenases of the benzene/toluene/isopropylbenzene/biphenyl branch. Thus, the degradation of benzene via a dioxygenolytic pathway necessitates the recruitment of a functional 4-oxalocrotonate branch from elsewhere on the genome.

Figure 6: Gene cluster organization. Genes encoding enzymes catalysing upper pathways of aromatic degradation are highlighted in blue and the lower pathways in pink. Black shaped arrows represent key genes in the pathway. Enzyme superfamily and cluster they belong to are indicated below the gene organization. A-Isopropylbenzene operon from *P. putida* RE204; B- Naphthalene/salicylate operon from *P. putida* G7; C- Phenol operon from *P. putida* MT4; D- Tol pathway operon from pWW0 plasmid from *P. putida* mt-2; E- Catechol intradiol and F- Protocatechuate pathway operons from *P. putida* KT2440.

Soluble Diiron Monooxygenases – (Ring hydroxylating monooxygenases, RHMO)

The soluble diiron monooxygenases are an evolutionary related superfamily of enzymes able to initiate the degradation of benzene and toluene to phenol and methylphenol, and further to catechol (Leahy et al., 2003). These enzyme complexes are composed of a terminal heteromultimeric oxygenase formed by two or three subunits, and an electron transport system comprising a reductase and in some cases a ferredoxin (Leahy et al., 2003). This enzyme family can be divided in three clusters (Vilchez-Vargas et al., 2013a), where cluster-1

represents phenol/methylphenol monooxygenases, cluster-2 represents alkene monooxygenases and cluster-3 represents benzene/toluene monooxygenases (Figure 7).

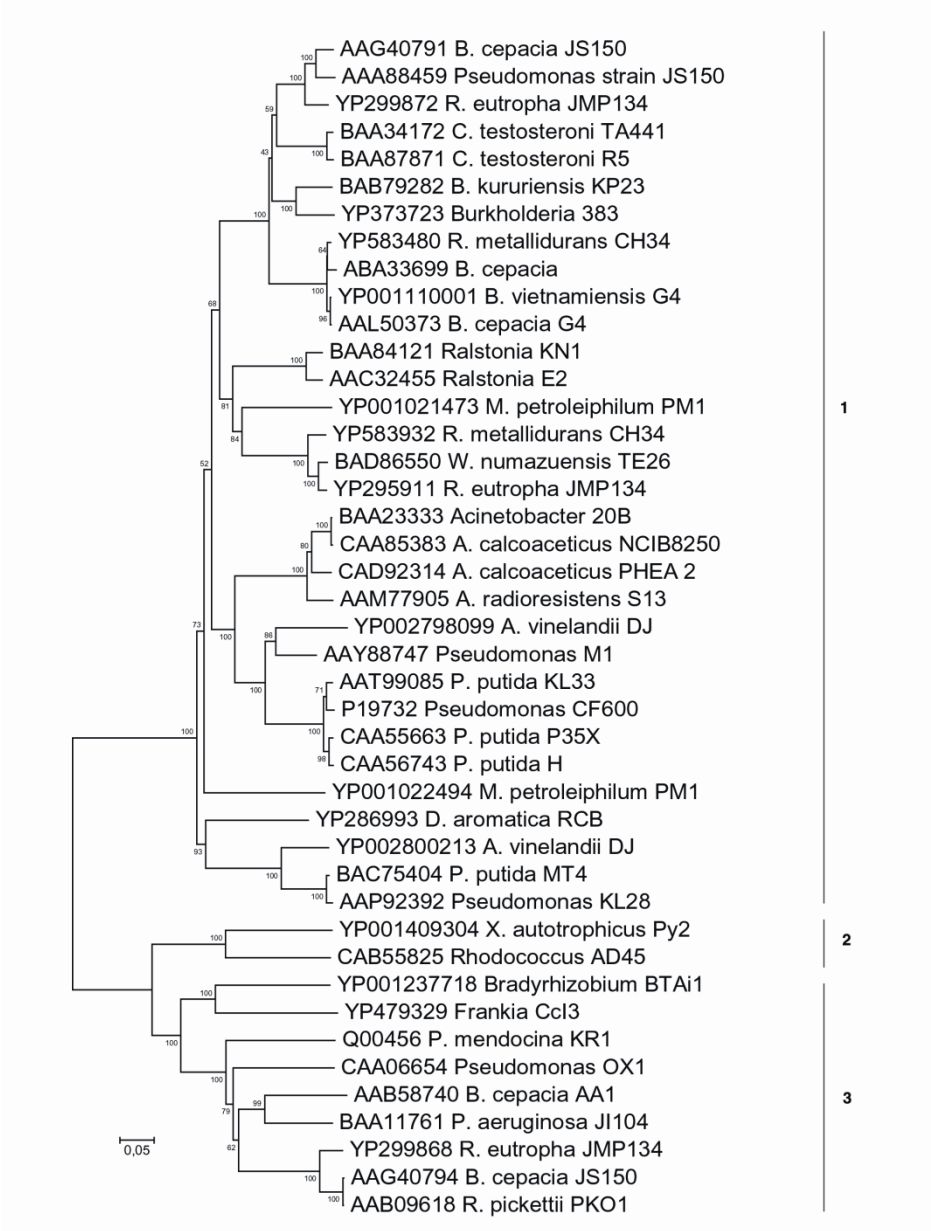


Figure 7: Evolutionary relationships among soluble diiron monooxygenases (RHMO). Cluster-1, phenol/methylphenol monooxygenases; cluster-2, alkene monooxygenases; cluster-3, benzene/toluene monooxygenases.

Enzymes in cluster-1 share the capability to hydroxylate phenol. Although phenol monooxygenases in general don't initiate the degradation of benzene and toluene, *Alicycliphilus denitrificans* strains BC and K601 harbour only a phenol monooxygenase in the

genome and are benzene and toluene degraders, suggesting that phenol monooxygenase is responsible for both the hydroxylation of benzene (and/or toluene) to (methyl-) phenol and the subsequent hydroxylation of (methyl-) phenol to (methyl-)catechol (Oosterkamp et al., 2013). It should also be mentioned that phenol monooxygenase from *Pseudomonas stutzeri* OX1 transforms benzene and toluene to catechol and 3-methylcatechol (via phenol and 2-methylphenol), respectively (Cafaro et al., 2004). Other described members in this cluster include the phenol monooxygenases of *P. putida* CF600 (Nordlund et al., 1990) and *P. putida* MT-4 (Takeo et al., 2006) (Figure 6-C).

The enzymes of cluster-3 are involved in benzene and toluene transformation. Even though some of the enzymes of this subfamily, like the toluene monooxygenases of *P. stutzeri* OX1 (Bertoni et al., 1998), *R. pickettii* PK01 and *P. mendocina* KR1 are able to oxidize also phenol and 3 methylphenols (Tao et al., 2004), toluene/benzene are the preferred substrates. Enzymes belonging to this cluster comprise TbcA toluene/benzene 2-monooxygenase from *B. cepacia* JS150 (Olsen et al., 1997), TbhA benzene/toluene hydroxylase from *B. cepacia* AA1 (Ma & Herson, 2000), and TmoA toluene monooxygenase from *P. mendocina* KR1, (Whited & Gibson, 1991). Interestingly also the alkene monooxygenase from *Xanthobacter* sp strain Py2, affiliated to cluster-2, has been reported to hydroxylate benzene and toluene (Zhou et al., 1999).

Integral membrane bound alkane hydroxylase (ALKB)

Integral membrane bound alkane hydroxylases catalyze a major step in the degradation of alkanes (Beilen & Funhoff, 2007). The p-cymene and xylene monooxygenases, a group of enzymes that catalyze the monooxygenation of the methylsubstituent of toluene, xylene isomers and p-cymene (Defrank & Ribbons, 1977; Harayama et al., 1989), also belong phylogenetically to this enzyme family.

The well known Tol pathway, mainly studied in *P. putida* mt-2 (Williams & Murray, 1974; Greated et al., 2002), initiates the degradation of toluene, 1,2,4-trimethylbenzene and of *m*- and *p*-xylene by a xylene monooxygenase (Figure 6-D) oxidizing the methylsubstituent to the alcohol. The channeling into the Krebs cycle is catalyzed by enzymes encoded in two

operons, where the *xylIUWCMABN* gene products are responsible for oxidizing the substrate to the respective benzoate (Harayama et al., 1986, 1989) and the *xylXYZLTEGFJQKIH* for the transformation of benzoate to catechol and further (Harayama & Rekik, 1990). Specifically, the *xyl*XYZ genes code for a broad substrate spectrum benzoate dioxygenase (Figure 5, cluster-13) and the *xyl*E gene codes for a catechol 2,3-dioxygenase, which is a member of the vicinal oxygen chelate superfamily of extradiol dioxygenases described later.

Extradiol dioxygenases of the vicinal oxygen chelate superfamily- EXDO

Extradiol dioxygenases of the vicinal oxygen chelate superfamily comprise enzymes that catalyze the dioxygenolytic ring fission of the catecholic intermediates in mono- and polyaromatic degradation (Vaillancourt et al., 2006). Phylogenetic analysis revealed that they can be divided into 13 clusters (Figure 8), where members of 5 of these clusters, namely EXDO-A; EXDO-D; EXDO-B; EXDO-K1 and EXDO-K2 have often been described to be involved in BTEX degradation. Genes encoding EXDO-D and EXDO-B extradiol dioxygenases are frequently found to cluster with genes encoding monooxygenases of the soluble diiron superfamily. For example a gene encoding an extradiol dioxygenase of the EXDO-D branch (Reut_B5687) is localized directly downstream of genes encoding a phenol hydroxylase (Reut_B5681-5686) in *C. necator* JMP 134 (Pérez-Pantoja et al., 2008) and the *bupB* gene (BAC75400) encoding an extradiol dioxygenase of the EXDO-B branch is localized upstream of the genes encoding butylphenol hydroxylase (*bup*A1A2A3A4A5A6) (BAC75401-BAC75406) in *P. putida* MT4 (Takeo et al., 2006) (Figure 6-C). EXDO-K1 and EXDO-K2 encoding genes are typically found in the same gene cluster with genes encoding benzene/toluene/isopropylbenzene/biphenyl dioxygenases. For example in the *ipb* operon of *P. putida* RE204, the *ipbC* gene (AAC03441) encoding an extradiol dioxygenase of the EXDO-K2 branch is localized in the same gene cluster as the genes encoding isopropylbenzene dioxygenase (*ipb*AaAbAcAd) (AAC03436-AAC03439) (Eaton et al., 1998) (Figure 6-A). Extradiol dioxygenases of cluster EXDO-A were observed to be involved in various different pathways, such as the Tol pathway described above, where toluene degradation is initiated by oxidation of the methylsubstituent (XylE, NP542866) (Williams & Murray, 1974; Greated et al., 2002) (Figure 6-D), the salicylate pathway of *P. putida* G7,

(NahH, YP534833) (Coitinho et al., 2012) (Figure 6-B) or the phenol degradation pathway of *Pseudomonas sp*. CF600 (DmpB, P17262) (Bartilson et al., 1990).

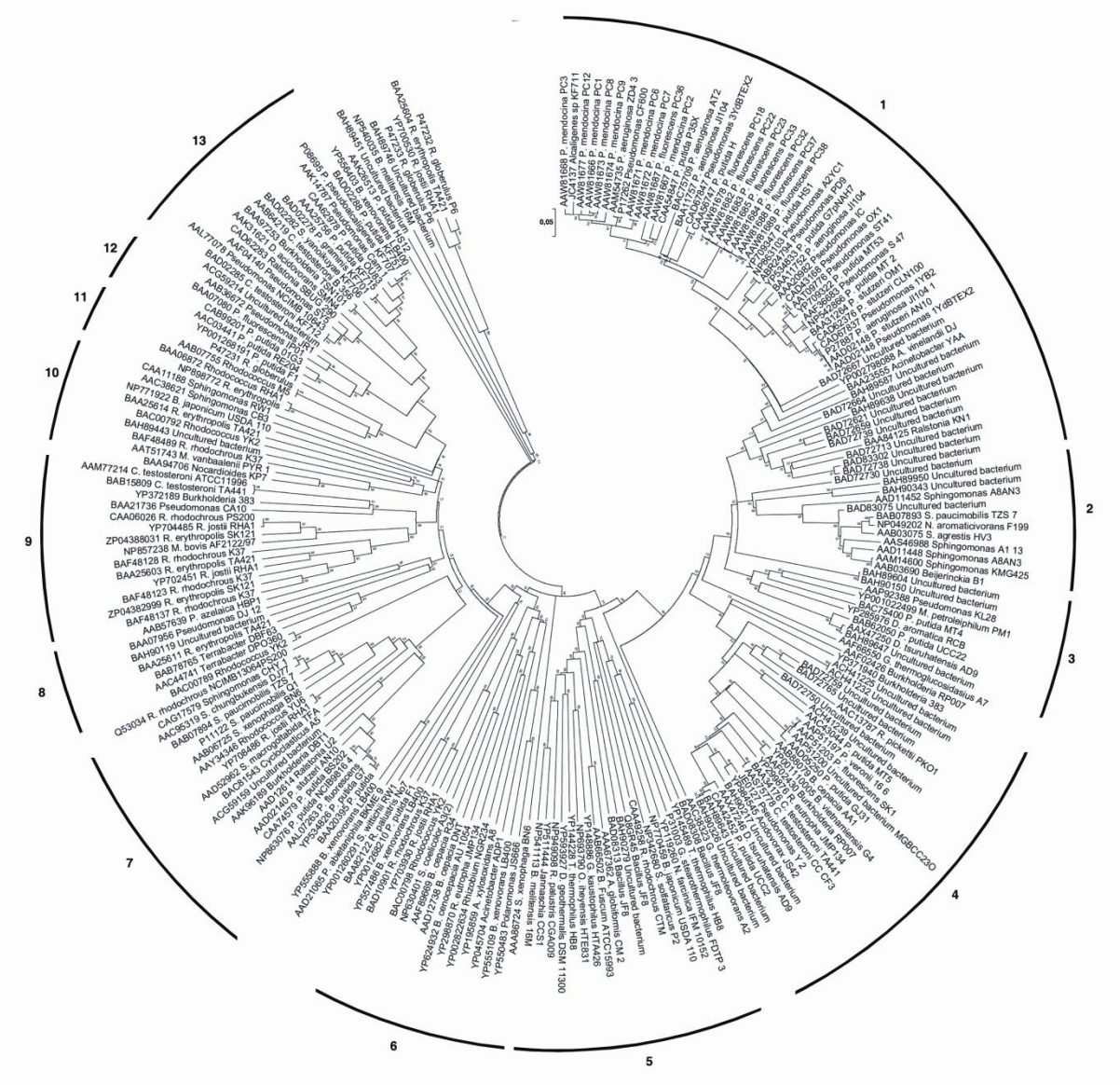


Figure 8: Evolutionary relationships of extradiol dioxygenases of the vicinal oxygenchelate superfamily (EXDO). Cluster-1, subfamily EXDO-A; cluster-2, subfamily EXDO-C; cluster-3, subfamily EXDO-B; cluster-4, subfamily EXDO-D; cluster-5, subfamily EXDO-E; cluster-6, subfamily EXDO-F; cluster-7, subfamily EXDO-G; cluster-8, subfamily EXDO-H; cluster-9, subfamily EXDO-I; cluster-10, subfamily EXDO-J; cluster-11, subfamily EXDO-L; cluster-12, subfamily EXDO-K2 and cluster-13, subfamily EXDO-K1.

Intradiol dioxygenases – INDO

Intradiol dioxygenases constitute a superfamily of enzymes that cleave catechol, protocatechuate or hydroxybenzoquinol between the hydroxyl groups. This superfamily can be divided into 6 subfamilies (Figure 9).

In *Pseudomonas* strains such as *P. putida* KT2440 or *Pseudomonas aeruginosa* PAO1, the enzymes encoded by the *cat*ABC operon (Figure 6-E) catalyze the cleavage of catechol to *cis,cis*-muconate, that is transformed to muconolactone and isomerized to the respective enol-lactone. Further metabolism then proceeds via 3-oxoadipate to Krebs cycle intermediates (Harwood & Parales, 1996). The enzymes encoded by the protocatechuate operon (*pca*) (Figure 6-F) from KT2440 and PAO1 cleave protocatechuate to 3-carboxy-cis,cis-muconate, that is further transformed to 4-carboxymuconolactone and then decarboxylated to the enol-lactone. Catechol and protocatechuate pathways thus merge at the step of enol-lactone formation. The further transformation to Krebs cycle intermediates in these strains (KT2440 and PAO1), is carried out by the *pca*IJF gene products (Jiménez et al., 2008). In hydroxyquinol degradation, ring-cleavage yields 3-hydroxy-cis,cis-muconate and a maleylacetate reductase forms 3-oxoadipate (Armengaud & Timmis, 1999; Perez-Pantoja et al., 2009).

Intradiol dioxygenase encoding genes are nearly exclusively found in the genomes of Actinobacteria and Proteobacteria. This contrasts the localization of extradiol dioxygenase encoding genes that are often encoded on plasmids (Pérez-Pantoja et al., 2012). Whereas, as described above, extradiol dioxygenases are typically involved in the degradation of BTEX compounds, the involvement of intradiol dioxygenases in BTEX degradation has only seldom been reported and despite the widespread of intradiol dioxygenases, to our knowledge, *Pseudomonas putida* ML2 (Fong et al., 1996) and *Burkholderia fungorum* FLU100 (Strunk & Engesser, 2013) are the only strains reported thus far degrading benzene via an intradiol-cleavage pathway.

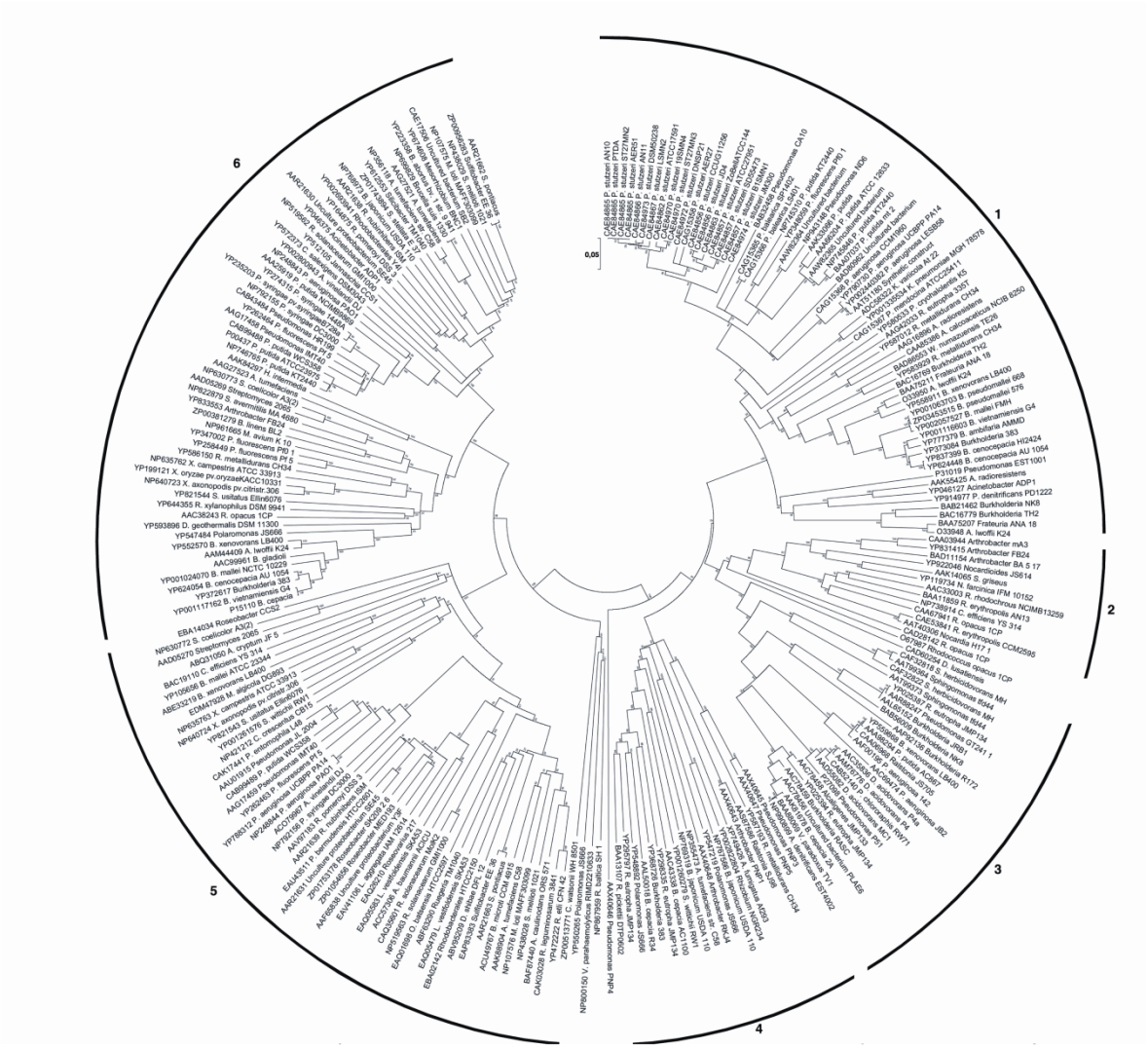


Figure 9: Evolutionary relationships among members of the Intradiol dioxygenase superfamily (INDO). Cluster-1, proteobacterial catechol 1,2-dioxygenases; cluster-2, actinobacterial catechol 1,2-dioxygenases; cluster-3, chlorocatechol 1,2-dioxygenases; cluster-4, hydroxybenzoquinol 1,2-dioxygenases; cluster-5, α-subunits of protocatechuate 3,4-dioxygenases; cluster-6, β-subunits of protocatechuate 3,4-dioxygenases.

1.3 Investigating isolates, communities and catabolic genes involved in biodegradation

The remediation by microorganisms is considered an efficient and cost effective method to clean-up contaminated sites (Atlas et al., 1991; Macnaughton et al., 1999; Samanta et al., 2002). By this process, the contaminants should be transformed to less toxic forms and finally mineralized. However other processes like sorption, volatilization, dispersion, dissolution and advection can occur, or pollutants may even be transformed into more toxic products, as may be the case during the anaerobic transformation of trichloroethylene (Vogel & McCarty, 1985; Tandoi et al., 1995), where vinylchloride is formed as an intermediate. Thus, the use of microorganisms to recover the environment requires detailed monitoring of the remediation process.

Molecular techniques have turned out to be especially amenable for detecting *in situ* pollutant-metabolizing microorganisms and catabolic key genes in a natural setting and have convincingly shown that xenobiotic degrading microorganisms enriched in the laboratory often don't play an important role in *in situ* biodegradation of pollutants (Jeon et al., 2003; Witzig et al., 2006). However, microorganisms important for biodegradation at selected sites can be identified and in some cases isolated, in this way permitting gaining valuable information on adaptation mechanisms.

1.3.1 Current molecular fingerprinting tools at our disposal

To gain an overview of bioremediation processes within the environment, it is important to identify potential key players, genes and pathways. This is difficult via cultivation-based techniques, as only a small percentage of the microorganisms from natural niches can be cultivated (Amann et al., 1995). Thus, culture-independent techniques targeting the bacterial 16S rRNA gene or targeting genes of specific function in biodegradation have been extensively used in the recent decades (Hugenholtz et al., 1998; Vilchez-Vargas et al., 2010).

One of the first culture-independent methods to study microbial communities was to create a clone library followed by Sanger sequencing (Sanger et al., 1973). In brief, purified

environmental DNA or PCR products were cloned into appropriate vectors, transformed into competent bacteria, amplified with specific vector primers and sequenced, permitting specific gene and species identification based on similarity to known sequences present in sequence databases. Clone libraries offer a precise method to study community individuals and catabolic genes, but are costly and laborious as the sequence depth needed per sample is substantial.

Some other molecular methods offer higher throughput than clone libraries at a fraction of the price, such as Terminal restriction fragment length polymorphism (T-RFLP) (Liu et al., 1997) and Single-strand conformation polymorphism (SSCP) (Orita et al., 1989). T-RFLP is a rapid, high-throughput method that produces labeled-fragments of different sizes from PCR amplicons that may correspond to a particular taxa or group of taxa or gene, producing a fingerprint of DNA fragments that can be used to gain a general overview of the community pattern under changing environmental conditions. Vázquez et al. (2009) used T-RFLP to study the community dynamics during bioremediation in diesel contaminated Antarctic soil. Although widely used, this method does not give direct sequence information, preventing direct catabolic gene and taxonomical identification, and requires some optimization prior to assigning names to corresponding restriction fragments (Camarinha-Silva et al., 2012). In contrast, SSCP can provide sequence information after amplicons are separated on an electrophoresis gel and sequenced. SSCP has successfully been applied to analyze the diversity of Rieske non-heme iron oxygenases and catechol 2,3-dioxygenases of environmental samples (Junca & Pieper, 2004; Witzig et al., 2006). However, it is laborious and, thus, is not regarded as a high throughput method.

More recently, deep-sequencing platforms that offer high-throughput sequencing with a range of associated costs and benefits have been introduced. There are two main platforms currently used, the 454-pyrosequencing (Margulies et al., 2005) and the Illumina deep-sequencing (Bentley et al., 2008). Sequencing via Roche's 454-pyrosequencing produces 'pyrotags' that are longer than Illumina's 'iTags', but more costly (Degnan & Ochman, 2012).

Pyrosequencing involves the detection of the released pyrophosphate during DNA synthesis (Hyman, 1988; Ronaghi, 1998). For pyrosequencing, the DNA is fragmented, ligated

to specific adapters and separated into single strands. These fragments bind to beads, which are captured in droplets and PCR amplified. The resulting double-stranded DNA is denatured and distributed into the wells of a fibre-optic slide and it is on this slide that the pyrophosphate sequencing take place (Margulies et al., 2005). The current 454/Roche **GS** FLX Titanium XL+ platform can generate 1 million sequence reads of up to 1kb length in a single run. However, it is expensive when compared with other technologies such as Illumina's instruments (Kircher & Kelso, 2010).

The Illumina method applies a reversible terminator technology. The first step in the procedure is the fragmentation of the genomic DNA that is then ligated to specific forked adapters to both ends of the DNA strand. Using specific primers targeting the adapter sequence, the formed library is amplified. This product is melted with sodium hydroxide, creating single stranded DNA (ssDNA). The ssDNA is bound to a flow cell that contains oligonucleotides complementary to the adapter ends. Priming occurs when the free end of the ligated fragment "bridges" with a complementary oligo on the flow cell surface. The bridge amplification process continues to create clusters of around 1,000 copies of the original sequence. Sequencing is performed through a reversible terminator-based method that enables detection of single bases as they are incorporated into growing DNA strands. A fluorescently-labeled terminator is imaged as each dNTP is added and then cleaved permitting incorporation of the next base. This step is continuously repeated, making a series of images that are automatically collected. To resolve the sequence of every cluster, the fluorescent signal is quantified and a base-calling algorithm applied (Bentley et al., 2008). Currently the HiSeq 2500 gives 6 billion paired-end reads. Because of the in vitro amplification steps during the library and flow cell preparation the average error rate rises to values of 10^{-2} to 10^{-3} (Kircher & Kelso, 2010). On the other hand the 454 GS-FLX Titanium pyrosequencing has a error rate of 10^{-3} to 10^{-4} (Quinlan et al., 2008).

Next generation sequencing has on the one hand been applied for genome sequencing and several microbial genomes have been sequenced using 454-pyrosequencing, such as the genome of the nitrogen-fixing endophyte *Klebsiella pneumoniae* 342 (Fouts et al., 2008), of four Campylobacter strains (*C. coli* RM2228, *C. lari* RM2100, *C. upsaliensis* RM3195, *C. jejuni* NCTC 11168 and RM1221) (Fouts et al., 2005), *Corynebacterium*

urealyticum (Tauch et al., 2008), *Pseudomonas extremaustralis* (Tribelli et al., 2012) and many others. Previously, the Illumina platform has mainly been used for re-sequencing genomes, aiming to detect small mutations in similar genomes, like for example in a study on *Acinetobacter baumannii* where the genome before and after therapy with tigecycline was analyzed (Hornsey et al., 2011).

The recent updates in Illumina technology, made it possible to sequence genomes exclusively by this method. The genome from *Pseudomonas syringae* pathovar *tabaci* 11528 was sequenced using exclusively Illumina technology. The assembly resulted in 71 contigs (Studholme et al., 2009), whereas the *Mycobacterium massiliense* Strain M154 genome sequencing project resulted in 44 contigs (Choo et al., 2012), and the *Bradyrhizobium elkanii* 587 sequencing project resulted in 260 scaffolds (series of contigs that are in the right order but not necessarily connected in one continuous sequence) (de Souza et al., 2012). Recently, two strains of *Legionella dumoffii* were sequenced using only Illumina and the assembly resulted in only 3 scaffolds, for the chromosome and 2 plasmids (Qin et al., 2012).

Another alternative largely used now a days is to integrate the output from the Illumina and pyrosequencing platforms facilitating gap closures as shown for the genome sequencing projects of *Geobacter sulfurreducens* KN400 (Nagarajan et al., 2010) *Paenibacillus vortex* (Sirota-Madi et al., 2010), *Yersina pestis* KIM D27 (Losada et al., 2011), *Yoka poxvirus* (Zhao et al., 2011) and *Pneumocystis jirovecii* (Cissé et al., 2012) among others. The hybrid technique can generate a better assembly, but in general does not result in genome closure. However for a good overview of a genome, when the target is the understanding of the physiology of a microorganism, a completely closed genome is not necessary, specially because gaps are usually localized in sequence repeats, regions representing the greatest challenge to all assembly algorithms (Phillippy et al., 2008).

The demand to process more samples with greater coverage at lower costs, trough high throughput technology, has led to developments for multiplexing samples in both Illumina and pyrosequencing platforms. This is specifically needed in the case of determining the community structure, where 16S rRNA amplicons are used. For this purpose barcodes were incorporated in Illumina's adapter sequence (Cronn et al., 2008; Corneveaux et al., 2011) and integration of the adapters and barcodes into a PCR primer (Caporaso et al., 2010,

Bartram et al., 2011, Zhou et al., 2011) significantly increased the number of samples that can be simultaneously pooled and analyzed per run.

A handful of publications have now assessed the applicability of using Illumina for amplicon deep-sequencing, and reported its advantages and shortcomings over other technologies, particularly in relation to error-rates, the need for quality filtering and taxonomic classification (Caporaso et al., 2010; Degnan & Ochman, 2012; Soergel et al., 2012; Werner et al., 2012; Bokulich et al., 2013). Even though there is a fear that short read lengths can affect classification quality, it has been shown that reads of 100-200 bases can yield a similar clustering pattern as longer 16S rRNA gene sequences (Liu et al., 2007). Moreover, taxonomic assignment can be very accurate at the order level and in some cases at the genus level (Lazarevic et al., 2009; Bartram et al., 2011; Caporaso et al., 2010; Claesson et al., 2010).

Microarrays are a powerful genomic technology that generates massive information from single organisms and environmental samples. This technology has been widely applied to study gene expression in pure cultures. To do this, the whole genome complement of a target organism is printed on a slide. This technique has then been used to analyze the response of *Sphingomonas wittichii* RW1 to dibenzofuran or salicylate exposure (Coronado et al., 2012), the gene expression of *Ferroglobus placidus* during the anaerobic metabolism of benzoate and phenol at high temperatures (Holmes et al., 2012), the gene expression of *Burkholderia xenovorans* LB400, a well-known polychlorinated biphenyl degrading strain (Patrauchan et al., 2011) during phenylacetate catabolism or the reactions of *Nitrosomonas europaea* biofilms when exposed to phenol and toluene (Lauchnor et al., 2011).

Among the large number of environmental oligoarray applications, the so called functional gene arrays are employed to measure genes involved in important environmental processes. First reported in 2007 the GeoChip had 24,243 oligonucleotide probes, covering > 10,000 genes from > 150 functional groups involved in nitrogen, carbon, sulfur and phosphorus cycling, metal reduction and resistance and organic contaminant degradation (He et al., 2007). In 2009 a second version (GeoChip 2.0), was published (Berthrong et al., 2009) and in 2010 the GeoChip 3.0 came up (He et al., 2010). The GeoChip was used to analyze catabolic genes (DNA) before and after crude oil degradation (Liang et al., 2009);

arsenic contamination (Xiong et al., 2012), in phenanthrene-spiked soil (Ding et al., 2012), an uranium-contaminated aquifer (Van Nostrand et al., 2011), to analyze nitrogen fixation genes in soils (Yergeau et al., 2007) and during diverse processes in different environments like mangroves (Bai et al., 2013) wetlands (Wang et al., 2012) rhizosphere (Trivedi et al., 2012) among others. Even though reports on the application specifically of the different versions of the GeoChip are overwhelming (He et al., 2007, 2010; Wang et al., 2009; Van Nostrand et al., 2011), the Geochip has severe shortcomings.

A comparison of methods used in the various recently published Geochip manuscripts (Vilchez-Vargas et al., 2013) shows that, even though the same array was used, methods differ significantly between different publications, specifically concerning the hybridization temperature, but also regarding buffer concentrations or washing conditions, all of which influence the hybridization outcome. Even more striking is the fact that even though both publications describing the first time the GeoChip2 and GeoChip3 versions, clearly show in figures that an immense amount of false positive signals is achieved when using hybridization temperatures below 45°C (only tested with synthesized oligonucleotides), and even though this fact is explicitly mentioned, environmental applications by that group typically use 42°C, with temperatures down to 35°C, such that it is clear that a significant proportion of observed signals will be false positive. However, the amount of false negative/positive signals is never clearly mentioned, but hidden behind miscalculations.

Moreover it is essential to upgrade the databases on aromatic degradation genes, in order to reliably correlate a hybridization signal to an underlying activity (Perez-Pantoja et al., 2009). Such a work is important given that the level of misannotation was identified as being as much as 80% for some subfamilies (Brennerova et al., 2009; Schnoes et al., 2009). Also regarding proteins associated with biodegradation high levels of misannotations are present in public databases (Perez-Pantoja et al., 2009; Vilchez-Vargas et al., 2010). As an example, various recent publications applying one of the versions of the GeoChip report on the presence of genes indicating aromatic degradation potential (Liang et al., 2009, 2011). However, a careful investigation revealed a significant amount of misannotations in proteins where probes were designed (Vilchez-Vargas et al., 2013). As an example, mistakes were found in 57 probes supposed to indicate a catechol degradative potential (Liang et al., 2009,

2011), where 18 of those target proteins were not involved in aromatic degradation and a further 14 probes are targeting proteins involved in chlorocatechol, hydroxybenzoquinol or homoprotocatechuate rather than in catechol degradation, concluding that only 25 out of the 57 probes may indicate catechol metabolic potential (Vilchez-Vargas et al., 2013).

To avoid misleading annotations, the catabolic array used here (see Vilchez-Vargas et al., 2013) is based on curated databases and covers genes encoding enzymes for key reactions of aromatic degradation (see below). The array comprises 1500 probes, where the design facilitates to assign signals to the respective protein subfamilies, therefore directly inferring function and substrate specificity.

1.4 Summary of techniques employed in this work

Genome sequencing

From previous work on a benzene contaminated site (Witzig et al., 2006) various isolates capable to mineralize benzene and toluene or benzene only are available. Two isolates designated *Pseudomonas sp*. strain 1YB2 and *Pseudomonas sp*. strain 1YdBTEX2 closely related to *Pseudomonas veronii*, as deduced by 16S rRNA gene sequencing, were chosen for further analysis. These strains are assumed to represent key catabolic players at the site under study and harbor a unique catabolic gene organization for benzene and/or toluene degradation. The challenge in sequencing genomes using only Illumina technology is to assemble the highly fragmented output. In the present work we developed an assembly pipeline that allows sequencing bacterial genomes at low cost using only Illumina technology.

Microcosm experiments

The use of microcosms was employed to monitor the bacterial community structure, adaptations and catabolic potential of soil treated with benzene and BTEX (benzene, toluene, ethylbenzene and xylenes). For this purpose, contaminated soil samples from different geographical sites were collected and incubated in hermetically closed microcosms subjected to constant benzene and BTEX exposure. Samples were taken over time for community structure analysis using 16S Illumina amplicon sequencing and catabolic gene composition analysis using the catabolic microarray.

Amplicon sequencing

Taking advantage of the rapid improvements in sequencing technology, we developed a protocol to characterize the structure of microbial communities by Illumina paired-end sequencing in a high-throughput fashion. Although, this method will be described in more depth in chapter 4 of this thesis, in brief, DNA was extracted from microcosms and PCR was performed with primers targeting the V5/V6 hypervariable 16S rDNA regions. These primers integrate different barcodes to the different samples, such that after sequencing pooled samples, the origin of each sequence can be identified.

Catabolic Microarray

A catabolic microarray, developed within our group, targets catabolic key genes of pollutant degradation and gives rapid molecular diagnostics of samples. It allows identifying catabolic genes that are involved in the catabolism of contaminants in pure cultures but also in complex environmental samples. The array contains approximately 1500 probes designed from knowledge-based curated databases for key aromatic catabolic gene families and key alkane degradation genes. This design facilitates to assign positive signals to the respective protein subfamilies, therefore directly inferring function and substrate specificity. The protein sequences of validly described members of the different catabolic protein families were collected from reference reports and included Rieske non-heme iron ring hydroxylating (di)oxygenases (RHDO) (Gibson & Parales, 2000), extradiol dioxygenases of the vicinal chelate superfamily (EXDOI) (Bolin, 1996), intradiol dioxygenases (INDO) (Eulberg et al., 1998), soluble di-iron aromatic ring hydroxylating monooxygenases (RHMO) (Leahy et al., 2003), ferredoxins of multicomponent aromatic degradation enzymes (FERRE) (Nojiri et al., 2002), muconate cycloisomerases (MCIS), maleylacetate reductases (MACR), alkane hydroxylases of the integral membrane-bound monooxygenases (ALKB) (Van Beilen et al., 2003), Cytochrome P450 CYP153 alkane hydroxylases (CYP153) (Kubota et al., 2005), benzoyl coenzyme A reductases (BCOAR) (Song & Ward, 2005) and benzylsuccinate synthases (BSSA) (Winderl et al., 2007) (For more information see Vilchez-Vargas et al., 2013). The catabolic microarray enables us to analyze if significant changes of the catabolic gene landscape occur under different treatments of contaminated soils in the laboratory.

1.5 Research approach

While environmental spills with recalcitrant pollutants are a common problem worldwide, the reasons for some microorganisms persisting and successfully establishing under pollutant pressure is still to be elucidated. Studying complex community structures and their involvement in bioremediation is challenging and requires a multifaceted approach. The era of Next Generation Sequencing (NGS), in particular deep-sequencing via the Illumina platform and the development of microarray technologies has provided us with the necessary tools to be capable to study both particular species-of-interest and the communities as a whole, in depth and at low cost. Specifically in this work, I could; (1) genome sequence two key players (*P. veronii* 1YdBTEX2 and *P. veronii* 1YB2) known to be important for aromatic biodegradation; (2) phylogenetcially characterize the microbial community structure of important soils contaminated with aromatic compounds; and (3) elucidate the catabolic potential and behavior of soil communities under contaminant stress over time.

1.6 Specific aims:

- To sequence and assemble the genomes of two representative isolates *P. veronii* 1YdBTEX2 and *P. veronii* 1YB2 – Chapter 2 (This chapter only show *P. veronii* 1YdBTEX2 publication).

- To annotate and compare the genomes of *P. veronii* 1YdBTEX2 and *P. veronii* 1YB2, to elucidate the organization of the benzene and/or toluene catabolic pathways and to verify expression levels of key genes, to gain an overview on the metabolic properties and catabolic gene landscape of key players for bioremediation and to reveal why these isolates prevail in an environment under pollutant stress – Chapter 3.

- To analyze microbial community and catabolic gene structures of contaminated soils of different origins under constant and specific contamination with benzene and BTEX in microcosm experiments over time and to identify shifts in community structure and prevalence of key genes – Chapter 4.

1.7 References

Amann RI, Ludwig W, Schleifer KH, Amann RI & Ludwig W (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143-169.

Armengaud J & Timmis KN (1999) Degradative pathway gene cluster is linked to the Initial dibenzo-p-dioxin pathway genes in *Sphingomonas* sp. strain RW1. *J Bacteriol* **181**: 3452-3461.

Atlas R, Horowitz A, Krichevsky M & Bej A (1991) Response of microbial populations to environmental disturbance. *Microbial Ecol* **22***:* 249–256.

Bagley DM & Gossett JM (1990) Tetrachloroethene transformation to trichloroethene and cis-1,2-dichloroethene by sulfate-reducing enrichment cultures. *Appl Environ Microbiol* **56**: 2511–6.

Bai S, Li J, He Z, Nostrand J, Tian Y, Lin G, Zhou G & Zheng T (2013) GeoChip-based analysis of the functional gene diversity and metabolic potential of soil microbial communities of mangroves. *Appl Microbiol Biotechnol* **97***:* 7035–7048.

Bartilson M, Nordlund I & Shingler V (1990) Location and organization of the dimethylphenol catabolic genes of *Pseudomonas* CF600. *Mol Gen Genet* **220**: 249-300.

Bartram AK, Lynch MD, Stearns JC, Moreno-Hagelsieb G & Neufeld JD (2011) Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl Environ Microbiol* **77**: 3846-3852.

Beilen J & Funhoff E (2007) Alkane hydroxylases involved in microbial alkane degradation. *Appl Microbiol Biotechnol* **74***:* 13–21.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ,Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ,Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S,Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA,Burrows JC, Carter NP, Castillo N, Chiara E Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B,Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP,Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S,Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-

Gonzales PI, Kumar A,Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW,Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D,Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A,Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA,Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC,Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R & Smith AJ (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53–59.

Berthrong ST, Schadt CW, Piñeiro G & Jackson RB (2009) Afforestation alters the composition of functional genes in soil and biogeochemical processes in South American grasslands. *Appl Environ Microbiol* **75**: 6240–6248.

Bertin PN, Médigue C & Normand P (2008) Advances in environmental genomics: towards an integrated view of micro-organisms and ecosystems. *Microbiology* **154***:* 347–59.

Bertoni G, Martino M, Galli E & Barbieri P (1998) Analysis of the gene cluster encoding toluene/o-xylene monooxygenase from *Pseudomonas stutzeri* OX1. *Appl Environ Microbiol* **64**: 3626-3632.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods*, **10**: 57–59.

Bravo-Linares CM & Mudge SM (2007) Analysis of volatile organic compounds (VOCs) in sediments using in situ SPME sampling. *J Environ Monit* **9**: 411–418.

Brennerova MV, Josefiova J, Brenner V, Pieper DH & Junca H (2009) Metagenomics reveals diversity and abundance of *meta*-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environ Microbiol* **11**: 2216–2227.

Cafaro V, Izzo V, Scognamiglio R, Notomista E, Capasso P, Casbarra A, Pucci P &Di Donato A (2004) Phenol hydroxylase and toluene/o-xylene monooxygenase from *Pseudomonas stutzeri* OX1: Interplay between Two Enzymes. *Appl Environ Microbiol* **70**: 2211-2219.

Camarinha-Silva A, Wos-Oxley ML, Jáuregui R, Becker K & Pieper DH (2012) Validating T-RFLP as a sensitive and high-throughput approach to assess bacterial diversity patterns in human anterior nares. *FEMS Microbiol Ecol* **79**: 98–108.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G & Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**: 1621–1624.

Choo SW, Wong YL, Tan JL, Ong CS, Wong GJ, Ng KP & Ngeow YF (2012) Annotated genome sequence of *Mycobacterium massiliense* strain M154, belonging to the recently created taxon *Mycobacterium abscessus* subsp. bolletii comb. nov. *J Bacteriol* **194**: 4778.

Cissé OH, Pagni M & Hauser PM (2012) De novo assembly of the *Pneumocystis jirovecii* genome from a single bronchoalveolar lavage fluid specimen from a patient. *MBio* **4**: e00428–12.

Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP & O'Toole PW (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**: e200.

Coitinho JB, Costa DMA, Guimarães SL, de Góes A M & Nagem RAP (2012) Expression, purification and preliminary crystallographic studies of NahF, a salicylaldehyde dehydrogenase from *Pseudomonas putida* G7 involved in naphthalene degradation. *Acta Crystallogr* **68**: 93–97.

Coronado E, Roggo C, Johnson DR & van der Meer JR (2012) Genome-wide analysis of salicylate and dibenzofuran metabolism in *Sphingomonas Wittichii* RW1. *Front Microbiol* **3**: 300.

Correa SM, Arbilla G, Marques MRC & Oliveira KMPG (2012) The impact of BTEX emissions from gas stations into the atmosphere. *Atmos Pollut Res* **3**: 163–169.

Corseuil HX, Dal M & Marins M (1997) Contaminação de águas subterrâneas por derramamentos de gasolina: o problema é grave? *Enge Sanit Ambient* **2**: 50–54.

Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N & Huentelman MJ (2008) Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods* **5**: 887–893.

Cronn R, Liston A, Parks M, Gernandt DS, Shen R & Mockler T (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res* **36**: e122.

De Souza JAM, Tieppo, E, Magnani GDS, Alves,LMC, Cardoso RL, Cruz LM, De Oliveira LF, Raittz RT, Souza EM, Pedrosa FO & Lemos EGDM (2012) Draft genome sequence of the nitrogen-fixing symbiotic bacterium *Bradyrhizobium elkanii* 587. *J Bacteriol* **194**: 3547–3548.

Defrank J J & Ribbons DW (1977) p-Cymene pathway in *Pseudomonas putida*: initial reactions. *J Bacteriol* **129**: 1356-1364.

Degnan PH, & Ochman H (2012) Illumina-based analysis of microbial community diversity. *ISME J **6***: 183–194.

Ding G-C, Heuer H, He Z, Xie J, Zhou J & Smalla K (2012) More functional genes and convergent overall functional patterns detected by geochip in phenanthrene-spiked soils. *FEMS Microbiol Ecol* **82**: 148–156.

Dorn E & Knackmusst HJ (1978) Chemical structure and biodegradability of halogenated aromatic compounds. Substituent effects on 1,2-dioxygenation of catechol. *Biochem* J **335**: 85–94.

Eaton RW & Timmis KN (1986) Characterization of a plasmid- specified pathway for catabolism of isopropylbenzene in *Pseudomonas putida* RE204. *J Bacteriol* **168**: 123–131.

Eaton RW (1996) p-Cumate catabolic pathway in *Pseudomonas putida* Fl: cloning and characterization of DNA carrying the *cmt* operon. *J Bacteriol* **178**: 1351–1362.

Eaton RW, Selifonova OV & Gedney RM (1998) Isopropylbenzene catabolic pathway in *Pseudomonas putida* RE204: nucleotide sequence analysis of the *ipb* operon and neighboring DNA from pRE4. *Biodegradation **9***: 119–132.

Eltis LD & Bolin JT (1996) Evolutionary relationships among extradiol dioxygenases. *J Bacteriol.* **178**: 5930–5937.

Erickson BD & Mondello FJ (1992) Nucleotide sequencing and transcriptional mapping of the genes encoding biphenyl dioxygenase, a multicomponent polychlorinated-biphenyl-degrading enzyme in *Pseudomonas strain* LB400. *J Bacteriol* **174**: 2903–12.

Eulberg D, Kourbatova EM, Golovleva LA & Schlömann M (1998) Evolutionary relationship between chlorocatechol catabolic enzymes from *Rhodococcus opacus* 1CP and their counterparts in Proteobacteria: Sequence divergence and functional convergence. *J Bacteriol* **180**: 1082-1094.

Fishman A, Tao Y & Wood TK (2004) Toluene 3-monooxygenase of *Ralstonia pickettii* PKO1 Is a para-hydroxylating enzyme. *J Bacteriol* **186**: 3117-3123.

Fong KP, Goh CB & Tan HM (1996) Characterization and expression of the plasmid-borne bedD gene from *Pseudomonas putida* ML2, which codes for a NAD+-dependent cis-benzene dihydrodiol dehydrogenase. *J Bacteriol* **178**: 5592–5601.

Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, Ravel J, Brinkac LM, DeBoy RT, Parker CT, Daugherty SC, Dodson RJ, Durkin AS, Madupu R,Sullivan SA, Shetty JU, Ayodeji

MA, Shvartsbeyn A, Schatz MC, Badger JH, Fraser CM & Nelson KE (2005) Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. *PLoS Biol* **3**: e15.

Fouts DE, Tyler HL, DeBoy RT, Daugherty S, Ren Q, Badger JH, Durkin AS, Huot H, Shrivastava S, Kothari S, Dodson RJ, Mohamoud Y, Khouri H, Roesch LF,Krogfelt KA, Struve C, Triplett EW & Methé BA (2008) Complete genome sequence of the N2-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS Genet* **4**: e1000141.

Fujita EM, Campell DE, Zielinska B, Arnott WP & Chow JC (2011) Concentrations of air toxics in motor vehicle – Dominated environments. Research report 156. *Health Effects Institute,* Boston, MA.

Gibson DT, Koch JR & Kallio RE (1968a) Oxidative degradation of aromatic hydrocarbons by microorganisms. I. Enzymic formation of catechol from benzene. *Biochemistry **7**:* 2653–2662.

Gibson DT, Koch JR, Schuld CL & Kallio RE (1968b) Oxidative degradation of aromatic hydrocarbons by microorganisms. II. Metabolism of halogenated aromatic hydrocarbons. *Biochemistry **7***: 3795–3802.

Gibson DT & Parales RE (2000) Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Curr Opin Biotechnol* **11**: 236–43.

Greated A, Lambertsen L, Williams PA & Thomas CM (2002) Complete sequence of the IncP-9 TOL plasmid pWW0 from *Pseudomonas putida*. *Environ Microbiol* **4**: 856–871.

Harayama S, Leppik RA, Rekik M, Mermod N, Lehrbach PR & Timmis KN (1986) Gene order of the TOL catabolic plasmid upper pathway operon and oxidation of both toluene and benzyl alcohol by the *xyl*A product. *J Bacteriol* **167**: 455-461.

Harayama S, Mermod N, Rekik M, Lehrbach PR & Timmis KN (1987) Roles of the divergent branches of the *meta*-cleavage pathway in the degradation of benzoate and substitued benzoates. *J Bacteriol* **169**: 558-564.

Harayama S, Rekik M, Wubbolts M, Rose K, Leppik RA & Timmis KN (1989) Characterization of five genes in the upper-pathway operon of TOL plasmid pWW0 from *Pseudomonas putida* and identification of the gene products. *J Bacteriol* **171**: 5048–5055.

Harayama S & Rekik M (1990) The *meta* cleavage operon of TOL degradative plasmid pWWO comprises 13 genes. *Mol Gen Genet* **221**: 113–120.

Harwood CS & Parales RE (1996) The β-Ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol* **50**: 553–590.

Hayaishi, O. (1950). Pyrocatecase a new enzyme catalizing oxidative breakdown of pyrocatechin. *J Biochem* **37**: 371–374.

Hayaishi O, Rothbergs S, Mehler AH & Saito Y (1957) Studies on oxygenases; enzymatic formation of kynurenine from tryptophan. *J Biol Chem* **229**: 889-896.

He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL, Li X, Wu L, Gentry TJ, Yin Y, Liebich J, Hazen TC & Zhou J (2010) GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J* **4**: 1167–1179.

He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, Huang Z, Wu W, Gu B, Jardine P, Criddle C & Zhou J (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1**: 67–77.

Hendrickx B, Junca H, Vosahlova J, Lindner A, Rüegg I, Bucheli-Witschel M, Faber F, Egli T, Mau M, Schlömann M, Brennerova M, Brenner V, Pieper DH, Top EM, Dejonghe W, Bastiaens L & Springael D (2006) Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. *J Microbiolo Methods* ***64***: 250–265.

Holmes DE, Risso C, Smith JA & Lovley DR (2012) Genome-scale analysis of anaerobic benzoate and phenol metabolism in the hyperthermophilic archaeon *Ferroglobus placidus*. *ISME J* **6**: 146–157.

Hornsey M, Loman N, Wareham DW, Ellington MJ, Pallen MJ, Turton JF, Underwood A, Gaulton T, Thomas CP, Doumith M, Livermore DM & Woodford N (2011) Whole-genome comparison of two *Acinetobacter baumannii* isolates from a single patient, where resistance developed during tigecycline therapy. *J Antimicrob Chemother* **66**: 1499–1503.

Hugenholtz P, Goebel BM & Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**: 4765-4774.

Hyman ED (1988) A new method of sequencing DNA. *Anal Biochem* **174**: 423–436.

Jeon C O, Park W, Padmanabhan P, DeRito C, Snape JR & Madsen EL (2003) Discovery of a bacterium, with distinctive dioxygenase, that is responsible for in situ biodegradation in contaminated sediment. *Proc Natl Acad Sci U S A* **100**: 13591–13596.

Jiménez JI, Canales A, Jiménez-Barbero J, Ginalski K, Rychlewski L, García JL & Díaz E (2008) Deciphering the genetic determinants for aerobic nicotinic acid degradation: the *nic* cluster from *Pseudomonas putida* KT2440. *Proc Natl Acad Sci U S A* **105**: 11329-11334.

Junca H & Pieper DH (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. *Environ Microbiol* **6**: 95–110.

Kircher M & Kelso J (2010) High-throughput DNA sequencing concepts and limitations. *Bioessays* ***32:*** 524–536.

Kubota M, Nodate M, Yasumoto-Hirose M, Uchiyama T, Kagami O, Shizuri Y & Misawa N (2005) Isolation and functional analysis of cytochrome P450 CYP153A genes from various environments. *Biosci Biotechnol Biochemi* **69:** 2421–2430.

Lapertot M, Seignez C, Ebrahimi S, Delorme S & Peringer P (2007) Mass production of bacterial communities adapted to the degradation of volatile organic compounds (TEX). *Biodegradation* **18**: 343–350.

Lauchnor EG, Radniecki TS & Semprini L (2011) Inhibition and gene expression of *Nitrosomonas europaea* biofilms exposed to phenol and toluene. *Biotechnol Bioeng* **180**: 750–757.

Leahy JG, Batchelor PJ & Morcomb SM (2003) Evolution of the soluble diiron monooxygenases. *FEMS Microbiol Rev* **27**: 449–479.

Liang Y, Nostrand JD, Van Wang J, Zhang X, Zhou J & Li G (2009) Microarray-based functional gene analysis of soil microbial communities during ozonation and biodegradation of crude oil. *Chemosphere* **75**: 193–199.

Liang Y, Van Nostrand JD, Deng Y, He Z, Wu L, Zhang X, Li G & Zhou J (2011) Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *ISME J* **5**: 403–413.

Liu WT, Marsh TL, Cheng H & Forney LJ (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**: 4516–4522.

Losada L, Varga JJ, Hostetler J, Radune D, Kim M, Durkin S, Schneewind O & Nierman WC (2011) Genome sequencing and analysis of *Yersina pestis* KIM D27, an avirulent strain exempt from select agent regulation. *PloS One* **6**: e19054.

Ma Y & Herson DS (2000) The catechol 2,3-dioxygenase gene and toluene monooxygenase genes from *Burkholderia* sp. AA1, an isolate capable of degrading aliphatic hydrocarbons and toluene. *J Ind Microbiol Biotechnol* **25**: 127–131.

Macnaughton SJ, Stephen JR, Venosa AD, Davis GA, Chang Y & White DC (1999). Microbial population changes during bioremediation of an experimental oil spill. *Appl Environ Microbiol* **65**: 3566-3574.

Majumdar D (2011) BTEX in Ambient air of a metropolitan city. *Journal of Environmental Protection* **02**: 11–20.

Nordlund I, Powlowski J & Shingler V (1990) Complete nucleotide sequence and polypeptide analysis of multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600. *J Bacteriol* **172**: 6826-6833.

Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Osterås M, Schrenzel J & François P (2013) Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods* **79:** 266–271.

Mao, D, Van De Weghe H, Diels L, De Brucker N, Lookman, R & Vanermen G (2008) High-performance liquid chromatography fractionation using a silver-modified column followed by two-dimensional comprehensive gas chromatography for detailed group-type characterization of oils and oil pollutions. *J Chromatogr A* **1179**: 33–40.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV,Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M,Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF & Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.

Masai E, Yamada A, Healy JM, Hatta, T, Kimbara K, Fukuda M & Yano K (1995) Characterization of biphenyl catabolic genes of gram-positive polychlorinated biphenyl degrader *Rhodococcus* sp. strain RHA1. *Appl Environ Microbiol* **61**: 2079–85.

Murakami S, Kodama N, Shinke R & Aoki K (1997) Classification of catechol 1,2-dioxygenase family: sequence analysis of a gene for the catechol 1,2-dioxygenase showing high specificity for methylcatechols from Gram+ aniline-assimilating *Rhodococcus erythropolis* AN-13. *Gene* **185**: 49–54.

Nagarajan H, Butler JE, Klimes A, Qiu Y, Zengler K, Ward J, Young ND, Methé BA, Palsson BØ, Lovley DR & Barrett CL (2010) De Novo assembly of the complete genome of an enhanced electricity-producing variant of *Geobacter sulfurreducens* using only short reads. *PloS One* **5**: e10922.

Nojiri H, Kamakura M, Urata M, Tanaka T, Chung JS, Takemura T, Yoshida T, Habe H & Omori T (2002) Dioxin catabolic genes are dispersed on the *Terrabacter* sp. DBF63 genome. *Biochem Biophys Res Commun* **296**: 233–240.

Olsen RH, Kukor J.J & Kaphammer B (1994) A novel toluene-3-monooxygenase pathway cloned from *Pseudomonas pickettii* PKO1. *J Bacteriol* **176**: 3749–3756.

Olsen RH, Kukor JJ, Byrne AM & Johnson GR (1997) Evidence for the evolution of a single component phenol/cresol hydroxylase from a multicomponent toluene monooxygenase. *J Ind Microbiol Biotechnol* **19**: 360–368.

Oosterkamp MJ, Veuskens T, Talarico Saia F, Weelink SA, Goodwin LA, Daligault HE, Bruce DC, Detter JC, Tapia R, Han CS, Land ML, Hauser LJ,Langenhoff AA, Gerritse J, van Berkel WJ, Pieper DH, Junca H, Smidt H, Schraa G, Davids M, Schaap PJ, Plugge CM & Stams AJ (2013) Genome analysis and physiological comparison of *Alicycliphilus denitrificans* strains BC and K601(T.). *PloS One* **8**: e66971.

Orita M, Iwahana H, Kanazawa H, Hayashi K & Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A* **86**: 2766–2770.

Patrauchan M, Parnell JJ, McLeod M, Florizone C, Tiedje J & Eltis L (2011) Genomic analysis of the phenylacetyl-CoA pathway in *Burkholderia xenovorans* LB400. *Arch Microbioly* **193**: 641–650.

Pérez-Pantoja D, De la Iglesia R, Pieper DH & González B (2008) Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. *FEMS Microbiol Rev* **32**: 736–794.

Perez-Pantoja D, Donoso R, Junca H, Gonzalez B & Pieper DH (2009) Phylogenomics of aerobic bacterial degradation of aromatics. In Handbook of Hydrocarbon and Lipid Microbiology. Timmis KN (ed). Berlin Heidelberg: Springer-Verlag, 1356-1397.

Pérez-Pantoja D, Donoso R, Agulló L, Córdova M, Seeger M, Pieper DH & González B (2012). Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. *Environ Microbiol* **14**: 1091–1117.

Phillippy AM, Schatz MC & Pop M (2008) Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol* **9**: R55.

Pieper DH, Engesser K-H, Don RH, Timmis KN & Knackmuss H-J (1985) Modified *ortho*-cleavage pathway in *Alcaligenes eutrophus* JMP134 for the degradation of 4-methylcatechol. *FEMS Microbiol Lett* **29**: 63–67.

Qin T, Cui Y, Cen Z, Liang T, Ren H, Yang X, Shao Z & Song Y (2012) Draft genome sequences of two *Legionella dumoffii* strains, TEX-KL and NY-23. *J Bacteriol* **194**: 1251–1252.

Quinlan AR, Stewart DA, Stromberg MP & Marth GT (2008) Pyrobayes: an improved base caller for SNP discovery in pyrosequences. *Nat Meth* **5**: 179–181.

Ronaghi M (1998) DNA SEQUENCING: A Sequencing method based on Real-Time Pyrophosphate. *Science* **281**: 363–365.

Samanta U, Bahadur RP & Chakrabarti P (2002) Quantifying the accessible surface area of protein residues in their local environment. *Protein Eng* **15**: 659–667.

Sanger F, Donelson JE & Coulson AR (1973) Use of DNA Polymerase I primed by a synthetic oligonucleotide to determine a nucleotide sequence in phage fl DNA. *Proc Natl Acad Sci U S A* **70**: 1209–1213.

Schnatter AR, Rosamilia K & Wojcik NC (2005) Review of the literature on benzene exposure and leukemia subtypes. *Chem Biol Interac* **153-154**: 9–21.

Schnoes AM, Brown SD, Dodevski I & Babbitt PC (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* **5**: e1000605.

Shields MS, Montgomery SO, Chapman PJ, Cuskey SM & Pritchard PH (1989) Novel pathway of toluene catabolism in the trichloroethylene-degrading bacterium g4. *Appl Environ Microbiol* **55**: 1624–1629.

Shim H & Yang S-T (1999) Biodegradation of benzene, toluene, ethylbenzene, and o-xylene by a coculture of *Pseudomonas putida* and *Pseudomonas fluorescens* immobilized in a fibrous-bed bioreactor. *J Biotechnol* **67**: 99–112.

Shim H, Shin E & Yang S-T (2002) A continuous fibrous-bed bioreactor for BTEX biodegradation by a co-culture of *Pseudomonas putida* and *Pseudomonas fluorescens*. *Adv Environ Res* **7**: 203–216.

Sirota-Madi A, Olender T, Helman Y, Ingham C, Brainis I, Roth D, Hagi E, Brodsky L, Leshkowitz D, Galatenko V, Nikolaev V, Mugasimangalam RC,Bransburg-Zabary S, Gutnick DL, Lancet D & Ben-Jacob E (2010) Genome sequence of the pattern forming *Paenibacillus vortex* bacterium reveals potential for thriving in complex environments. *BMC Genomics*, **11**: 710.

Soergel DaW, Dey N, Knight R & Brenner SE (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J* **6**: 1440–1444.

Song B & Ward BB (2005) Genetic diversity of benzoyl coenzyme a reductase genes detected in denitrifying isolates and estuarine sediment communities. *Appl Environ Microbiol* **71**: 2036-2045.

Strunk N & Engesser K-H (2013) Degradation of fluorobenzene and its central metabolites 3-fluorocatechol and 2-fluoromuconate by *Burkholderia fungorum* FLU100. *Appl Microbiol Biotechnol* **97**: 5605–5614.

Studholme DJ, Ibanez SG, MacLean D, Dangl JL, Chang JH & Rathjen JP (2009) A draft genome sequence and functional screen reveals the repertoire of type III secreted proteins of *Pseudomonas syringae* pathovar tabaci 11528. *BMC Genomics* **10**: 395.

Takeo M, Prabu SK, Kitamura C, Hirai M, Takahashi H, Kato D-I & Negoro S (2006) Characterization of alkylphenol degradation gene cluster in *Pseudomonas putida* MT4 and evidence of oxidation of alkylphenols and alkylcatechols with medium-length alkyl chain. *J Biosci Bioeng* **102**: 352–361.

Tandoi V, Gossett JM & Maymo X (1995) Characterization of an H2 utilizing enrichment culture that reductively dechlorinates tetrachloroethene to vinyl chloride and ethene in the absence of methanogenesis and acetogenesis. *Appl Environ Microbiol* **61**: 3928–3933.

Tao,Y, Fishman A, Bentley WE, Wood K & Wood TK (2004) Oxidation of benzene to phenol, catechol, 4-monooxygenase of *Pseudomonas mendocina* KR1 and toluene 3-monooxygenase of *Ralstonia pickettii*. *Appl Environ Microbiol* **70**: 3814-3820.

Tauch A, Trost E, Tilker A, Ludewig U, Schneiker S, Goesmann A, Arnold W, Bekel T, Brinkrolf K, Brune I, Götker S, Kalinowski J, Kamp PB, Lobo FP, Viehoever P, Weisshaar B, Soriano F, Dröge M & Pühler A (2008) The lifestyle of *Corynebacterium urealyticum* derived from its complete genome sequence established by pyrosequencing. *J Biotechnol* **136**: 11–21.

Tribelli PM, Raiger Iustman LJ, Catone MV, Di Martino C, Revale S, Méndez BS & López NI (2012) Genome sequence of the polyhydroxybutyrate producer *Pseudomonas extremaustralis*, a highly stress-resistant Antarctic bacterium. *J Bacteriol* **194**: 2381–2382.

Trivedi P, He Z, Van Nostrand JD, Albrigo G, Zhou J & Wang N (2012) Huanglongbing alters the structure and functional diversity of microbial communities associated with citrus rhizosphere. *ISME J* **6**: 363–383.

Vaillancourt FH, Bolin JT & Eltis LD (2006) The ins and outs of ring-cleaving dioxygenases. *Crit Rev Biochem Mol Biol* **41**: 241–267.

Van Beilen JB, Mourlane F, Seeger MA, Kovac J, Li Z, Smits THM, Fritsche U & Witholt B (2003) Cloning of Baeyer-Villiger monooxygenases from Comamonas, Xanthobacter and Rhodococcus using polymerase chain reaction with highly degenerate primers. *Environ Microbiol* **5**: 174–182.

Van Nostrand JD, Wu L, Wu WM, Huang Z, Gentry TJ, Deng Y, Carley J, Carroll S, He Z, Gu B, Luo J, Criddle CS, Watson DB, Jardine PM, Marsh TL, Tiedje JM,Hazen TC & Zhou J (2011) Dynamics of microbial community composition and function during in situ bioremediation of a uranium-contaminated aquifer. *Appl Environ Microbiol* **77**: 3860–3869.

Vilchez-Vargas R, Junca H & Pieper DH (2010) Metabolic networks, Microbial Ecol and "omics" technologies: towards understanding in situ biodegradation processes. *Environ Microbiol* **12**: 3089–3104.

Vilchez-Vargas R, Geffers R, Suárez-Diez M, Conte I, Waliczek A, Kaser VS, Kralova M, Junca H & Pieper DH (2013) Analysis of the microbial gene landscape and transcriptome for aromatic pollutants and alkane degradation using a novel internally calibrated microarray system. *Environ Microbiol* **15**: 1016–1039.

Vogel TM & Mccarty PL (1985) Biotransformation of tetrachloroethylene to biotransformation of tetrachloroethylene to trichloroethylene, dichloroethylene, vinyl chloride, and carbon dioxide under methanogenic conditions. *Appl Environ Microbiol* **49**: 1080-1083.

Wang F, Zhou H, Meng J, Peng X, Jiang L, Sun P, Zhang C, Van Nostrand JD, Deng Y, He Z, Wu L, Zhou J & Xiao X (2009) GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *National Academy of Sciences (US)* **106**: 4840–4845.

Wang,H, He Z, Lu Z, Zhou J, Van Nostrand JD, Xu X & Zhang Z (2012) Genetic linkage of soil carbon pools and microbial functions in subtropical freshwater wetlands in response to experimental warming. *Appl Environ Microbiol* **78**: 7652–7661.

Werner JJ, Zhou D, Caporaso JG, Knight R & Angenent LT (2012) Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *ISME J* **6**: 1273–1276.

Whited GM & Gibson DT (1991) Separation and partial characterization of the enzymes of the toluene-4-monooxygenase catabolic pathway in *Pseudomonas mendocina* KR1. *J Bacteriol* **173**: 3017–3020.

Williams PA & Murray K (1974) Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida* (arvilla) mt-2: evidence for the existence of a TOL plasmid. *J Bacteriol* **120**: 416–423.

Winderl C, Schaefer S & Lueders T (2007) Detection of anaerobic toluene and hydrocarbon degraders in contaminated aquifers using benzylsuccinate synthase (*bss*A) genes as a functional marker. *Environ Microbiol* **9**: 1035–1046.

Witzig R, Junca H, Hecht HJ & Pieper DH (2006) Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: Links between benzene biodegradation and genes similar to those encoding isopropylbenzene dioxygenases. *Appl Environ Microbiol* **72**: 3504–3514.

Worsey MJ & Williams PA (1975) Metabolism of toluene and xylenes by *Pseudomonas putida* (arvilla) mt-2: evidence for a new function of the TOL plasmid. *J Bacteriol* **124**: 7-13.

Xiong, J, He Z, Van Nostrand JD, Luo G, Tu S, Zhou, J & Wang G (2012) Assessing the microbial community and functional genes in a vertical soil profile with long-term arsenic contamination. *PloS One* **7**: e50507.

Yen K-M & Gunsalus IC (1981) Plasmid gene organization: naphthalene/salicylate oxidation. *Proc Natl Acad Sci U S A* **79**: 874-878.

Yergeau E, Kang S, He Z, Zhou J & Kowalchuk GA (2007) Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *ISME J* **1**: 163–79.

Zhao G, Droit L, Tesh RB, Popov VL, Little NS, Upton C, Virgin HW & Wang D (2011) The genome of *Yoka poxvirus*. *J Virol* **85**: 10230–10238.

Zhou N, Jenkins A, Chan CKN, Chion K, Leak DJ & Chion KWO (1999) The alkene monooxygenase from *Xanthobacter* Strain Py2 is closely related to aromatic monooxygenases and catalyzes aromatic monohydroxylation benzene, toluene, and phenol. *Appl Environ Microbiol* **65**: 1589-1595.

Zielinski M, Backhaus S & Hofer B (2002) The principal determinants for the structure of the substrate-binding pocket are located within a central core of a biphenyl dioxygenase alpha subunit. *Microbiology* **148**: 2439–2448.

Chapter II – Genome Sequence of the soil bacterium *Pseudomonas veronii* strain 1YdBTEX2

Daiana de Lima-Morales[1], Diego Chaves Moreno[1], Michael Jarek[2], Ramiro Vilches-Vargas[1,3], Ruy Jauregui Sandoval[1], Dietmar H. Pieper[1]#

[1]Microbial Interactions and Processes Research Group, Department of Medical Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany; [2]Genom Analytics, Helmholtz Centre for Infection Research, Inhoffenstrasse 7, Braunschweig, Germany; [3]Present address: Laboratory of Microbial Ecology and Technology (LabMET), Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

Keywords: BTEX biodegradation, Illumina genome sequence, assembly Illumina output

Running title: A draft genome of *Pseudomonas veronii* 1YdBTEX2

Published in Genome Announcements

2.1 Abstract

Pseudomonas veronii 1YdBTEX2 was isolated from a benzene contaminated site. Here we report the draft genome of 1YdBTEX2 and its genes associated with aromatic metabolism. The broad catabolic potential of this strain is consistent with the environment in which it was isolated.

2.2 Genome Announcement

Environmental spills with recalcitrant pollutants are a common problem worldwide, threatening the environment and human health. The main mechanism for recovering of these sites is through biodegradation by indigenous microorganisms. The reasons for some microorganisms persisting and successfully establishing under pollutant pressure is still to be elucidated.

Here the draft genome of *Pseudomonas veronii* 1YdBTEX2 isolated from soil of an air-strip in the Czech Republic highly contaminated with benzene (Junca & Pieper, 2004) is presented. This strain harbors a unique catabolic pathway for the degradation of both benzene and toluene and has been purported to be a key player in the biodegradation of these pollutants at this site (Junca & Pieper, 2004; Witzig et al., 2006). While the 16S rRNA gene of 1YdBTEX2 is 99.6% similar to that of *P. veronii* CIP104663^T and 98.7% similar to that of *Pseudomonas fluorescens* IAM12022^T, the housekeeping gene gyrB is 99.2% similar to that of *P. veronii* CIP104663^T and only 91.0% similar to that of *P. fluorescens* IAM12022^T (http://www.uib.es/microbiologiaBD/Welcome.html). Thus, 1YdBTEX2 is the first member of the *P. veronii* species to be sequenced.

The genome was sequenced using the Illumina GAII genome analyzer generating 110 bp paired-end reads. Approximately 9.5 million reads were obtained and the whole-genome was assembled into 281 contigs using Velvet (Zerbino & Birney, 2008) and 344 contigs using Edena (Hernandez et al., 2008). Both datasets were then merged using Minimus2 (Sommer et al., 2007) resulting in 119 contigs. In order to merge contigs, scaffolds were constructed using SSPACE (Boetzer et al., 2011). Comparisons between contigs and scaffolds were then

made using Mauve (Darling et al., 2004), and compared to the genome of *P. fluorescens* SBW25 (Silby et al., 2004). Gaps between scaffolds were closed by PCR amplification followed by DNA Sanger sequencing. Annotation was performed with the RAST server (Rapid Annotations using Subsystems Technology) version 4.0 (Aziz et al., 2008) generating 5,981 candidate protein-encoding genes. The resulting genome sequence consists of 63 contigs (N_{50} = 229,690, L_{50} = 10) of 6,680,724 bp with a GC content of 59%. Besides the peripheral aromatic degradation pathways for benzene/toluene, phenol, salicylate, tryptophan (via anthranilate), ferulate (via vanillate), p-coumarate (via 4-hydroxybenzoate), 3-hydroxyphenylpropionate and phenylalanine (via 4-hydroxyphenylpyruvate), central aromatic degradation pathways for catechol (both via intra- and extradiol cleavage), protocatechuate (via intradiol cleavage), homogentisate and 2,3-dihydroxyphenylpropionate and an *alk*B gene encoding alkane monooxygenase were also observed in strain 1YdBTEX2. Furthermore, while 1YdBTEX2 is versatile with respect to its use of a range of carbon sources, a cluster of genes comprising a complete denitrification pathway including nitrate, nitrite, nitric oxide and nitrous oxide reduction was found, conferring flexibility in periods of anoxia. The presence of a (NiFe) hydrogenase where an entire operon similar to that found in *Ralstonia eutropha* H16 (Burgdorf et al., 2005) was identified here, which indicates that hydrogen may be used as an electron donor, a feature unknown to Pseudomonas strains with the exception of *P. extremaustralis* (Tribelli et al., 2012). The data is consistent with the environment from which 1YdBTEX2 was isolated and underlines its broad catabolic potential.

2.3 Nucleotide sequence accession numbers

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AOUH00000000. The version described in this paper is the first version, AOUH00000000.

2.4 Acknowledgements

This work was funded by BACSIN (project number: 211684) from the European Commission. Daiana de Lima Morales was supported by a grant from the CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (project number: 290020/2008-5). Special thanks to Dr. Robert Geffers from the sequencing facility at the Helmholtz Centre for Infection Research and to Iris Plumeier and Agnes Waliczek for technical assistance.

2.5 References

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A & Zagnitko O (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**:75.

Boetzer M, Henkel CV, Jansen HJ, Butler D & Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.

Burgdorf T, van der Linden E, Bernhard M, Yin QY, Back JW, Hartog AF, Muijsers AO, de Koster CG, Albracht SP & Friedrich B (2005) The soluble NAD+-Reducin [NiFe]-hydrogenase from *Ralstonia eutropha* H16 consists of six subunits and can be specifically activated by NADPH. *J Bacteriol* **187**: 3122-3132.

Darling A CE, Mau B, Blattner FR & Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394–403.

Hernandez D, François P, Farinelli L, Osterås M & Schrenzel J (2008) De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res* **18**: 802–809.

Junca H & Pieper DH (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. *Environ Microbiol* **6**: 95–110.

Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, Knight CG, Malone JG,Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS,Bentley SD, Hothersall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB & Thomson NR (2009) Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biol* **10**: R51.

Sommer DD, Delcher AL, Salzberg SL & Pop M (2007) Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* **8**: 64.

Tribelli PM, Raiger Iustman LJ, Catone MV, Di Martino C, Revale S, Méndez BS & López NI (2012) Genome sequence of the polyhydroxybutyrate producer *Pseudomonas extremaustralis*, a highly stress-resistant Antarctic bacterium. *J Bacteriol* **194**: 2381–2382.

Witzig R, Junca H, Hecht HJ & Pieper DH (2006) Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: Links between benzene biodegradation and genes similar to those encoding isopropylbenzene dioxygenases. *Appl Environ Microbiol* **72**: 3504–3514.

Zerbino DR & Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821–829.

Chapter III - Physiological diversity of *Pseudomonas veronii* 1YdBTEX2 and 1YB2, two versatile benzene degrading microorganisms

Daiana Lima-Morales, Diego Chaves-Moreno, Melissa L Wos-Oxley, Ruy Jáuregui, Ramiro Vilchez-Vargas*, Dietmar H. Pieper

Microbial Interactions and Processes Research Group, Department of Medical Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany; *Present address Laboratory of Microbial Ecology and Technology (LabMET), University of Ghent, Ghent, Belgium

Keywords: benzene, Illumina, *P. veronii*, denitrification, [NiFe]-hydrogenase

Running title: *Pseudomonas veronii* 1YdBTEX2 and 1YB2 genome

Physiological diversity of *Pseudomonas veronii* 1YdBTEX2 and 1YB2, two versatile benzene degrading microorganisms

3.1 Abstract

Comparison of the genomes of strains *Pseudomonas veronii* 1YdBTEX2 and 1YB2 with those of previously sequenced strains indicated that they belong to a species different from *Pseudomonas fluorescens*, the high similarity of the 16S rDNA and the *gyrB* gene to those of the type strain indicated that they are in fact *P. veronii* isolates. Analysis revealed that they harbor unique catabolic pathways for the degradation of benzene. These pathways comprise a gene cluster encoding an isopropylbenzene dioxygenase where genes encoding downstream enzymes were interrupted by stop codons. Extradiol dioxygenases were recruited from gene clusters comprising genes encoding a 2-hydroxymuconic semialdehyde dehydrogenase necessary for benzene degradation but typically absent from isopropylbenzene dioxygenase encoding gene clusters. The benzene dihydrodiol dehydrogenase encoding gene was not clustered with any other aromatic degradation gene and the encoded protein was only distantly related to dehydrogenases of aromatic degradation pathways. The involvement of the different gene clusters in the degradation pathways was verified through RT-PCR. Both *Pseudomonas veronii* strains harbour unique features as gene clusters for the synthesis of the siderophore pyochelin, a complete denitrification pathway organized in one superoperonic cluster and a gene cluster comprising a [NiFe]-hydrogenase for the use of hydrogen as alternative electron donor. These genetic attributes may help *P. veronii* 1YdBTEX2 and 1YB2 to survive under environmental conditions.

3.2 Introduction

Members of the *Pseudomonas* genus are highly versatile and are known to colonize diverse habitats such as humans, animals, plants, water and soil (Silby et al., 2011). Thus far, the genomes of hundreds of *Pseudomonas* isolates of various species have been sequenced (Silby et al., 2011) and in accordance to the large metabolic diversity of this genus, only 25% to 35% of the genome of each strain is composed of core genes shared by all members of the genus (Loper et al., 2012). More than 200 *Pseudomonas* species have been named so far (www.bacterio.net/pseudomonas.html). Based on the detailed analysis of 107 *Pseudomonas* species, Mulet et al. (2010) grouped these strains into intrageneric groups and subgroups, where the *Pseudomonas fluorescens* subgroup comprised 20 species. *P. veronii,* one of the species of the *P. fluorescens* subgroup was first isolated in France from natural mineral water (Elomari et al. 1996). *P. veronii* strains have since then been implicated in human pathogenicity (Cheuk et al. 2000), heavy metal biosorption (Vullo et al., 2008), dibenzo-p-dioxin degradation (Hong et al., 2004), pentachlorophenol transformation (Nam et al., 2003) and chloroaromatic and alkyl methyl ketones degradation (Hong et al. 2004; Onaca et al. 2007).

P. veronii strains were also pointed out as key catabolic players in a BTEX contaminated area where benzene had been the major contaminant (Junca & Pieper, 2004). Two strains, termed *P. veronii* 1YB2 and *P. veronii* 1YdBTEX2 had been subjected to a detailed analysis, as they contained catabolic gene variants shown to be highly abundant at the site. Interestingly, both strains differed in their substrate specificity, and whereas *P. veronii* 1YdBTEX2 grows on benzene and toluene, strain 1YB2 grows on benzene only (Junca & Pieper, 2004; Witzig et al., 2006). The difference in substrate specificity was due to small sequence differences in the α-subunit of the initial multicomponent Rieske non-heme iron oxygenase responsible for activation of the aromatic ring. Even though both oxygenases showed highest similarity to isopropylbenzene dioxygenases, isolates harbouring an oxygenase carrying two methionine residues that narrow the available space in the substrate-binding pocket could only attack and degrade benzene but not toluene.

The α-subunit sequences of the Rieske non-heme iron oxygenases cluster into families associated with the substrates oxidized by the members forming this superfamily (Gibson & Parales, 2000) and, thus, enzymes involved in the degradation of benzene and toluene are expected to show highest similarities to those of the benzene/toluene dioxygenase cluster. However the strains under study here harbor dioxygenases most closely related to isopropylbenzene dioxygenases (Witzig et al., 2006). In the well-characterized isopropylbenzene degradative pathways of *Pseudomonas putida* RE204 and *Pseudomonas fluorescens* IP01 (Aoki et al., 1996; Eaton et al., 1996), the genes coding for the oxygenase subunits are organized in operons together with genes encoding a *cis*-dihydrodiol dehydrogenase and an extradiol dioxygenase of the subfamily I.3.A (Eltis et al., 1996). However, *P. veronii* 1YB2 and 1YdBTEX2 do not express a subfamily I.3.A catechol 2,3-dioxygenase during growth on benzene but recruit a subfamily I.2.A C23O for catechol cleavage (Junca et al., 2004). Such a recruitment of a subfamily I.2.A C23O has, to our knowledge, previously not been reported for the degradation of benzene or toluene via a dioxygenolytic route.

Since *P. veronii* 1YdBTEX2 and 1YB2 had been indicated to be catabolic key players with unique pathway arrangements, we sequenced the genome of both strains to elucidate the organization of the benzene and/or toluene catabolic pathways, to gain an overview on their metabolic properties and catabolic gene landscapes and to reveal why these isolates prevail in an environment under pollutant stress.

3.3 Material and Methods

3.3.1 Strains and culture conditions

P. veronii 1YB2 and *P. veronii* 1YdBTEX2 were cultured in mineral medium (Dorn et al., 1974) containing 50 mM phosphate buffer (pH 7.4) supplemented with trace elements. Benzene or toluene were supplied via the gas phase (2-5 mM each) (Junca & Pieper, 2004) and cultures incubated at room temperature without agitation. Growth experiments with different carbon sources were conducted in minimal medium

supplemented with ethylbenzene (2-5 mM, via the gas phase), phenol (5 mM), salicylate (5 mM) or anthranilate (2 mM).

3.3.2 Genome sequencing and assembling

Genomes were assembled according to Lima-Morales et al., (2013). Briefly both genomes were sequenced using the high-throughput Solexa sequencing technology on an Illumina GAII genome analyzer generating 110 bp paired-ends. Approximately 10 million short sequence reads were obtained for each genome. Whole-genome shotgun sequencing data for strains *P. veronii* 1YdBTEX2 and 1YB2 were assembled into 281 and 376 contigs respectively using Velvet (Zerbino & Birney**,** 2008), and 344 and 320 contigs respectively, using Edena (Hernandez et al., 2008). Both datasets for each strain were then merged using Minimus (Sommer et al., 2007), resulting in 119 contigs for strain 1YdBTEX2 and 159 contigs for strain 1YB2. In order to further diminish the number of gaps, contigs were rearranged by means of two strategies; firstly, scaffolds were constructed using SSPACE (Boetzer et al., 2011) and comparisons between contigs and scaffolds were made using MAUVE (Darling et al., 2004); secondly, contigs were compared to the *Pseudomonas fluorescens* SBW25 genome (Silby et al., 2009), the most similar genome according to the SEED data base (Aziz et al., 2005). These strategies resulted in a possible assembly for a number of contigs. In order to confirm this information and close gaps between contigs, primers framing those gaps were designed and PCR products sequenced with Sanger sequencing. Finally 63 and 115 contigs were obtained for strain *P. veronii* 1YdBTEX2 and 1YB2, respectively.

3.3.3 Phylogenetic comparisons

Nucleotide sequences of 16S rDNA and of DNA gyrase subunit B (*gyr*B), were aligned using MUSCLE 3.7 (Edgar, 2004). Phylogenetic trees were constructed with MEGA5 (Tamura et al., 2011) using the Neighbour-Joining (N-J) algorithm (Saitou & Nei, 1987) with Jukes-Cantor correction and pairwise deletion of gaps and missing data. A total of 100 bootstrap replications were done to test for branch robustness.

Homologous proteins were detected as described by Moreno-Hagelsieb & Latimer (2008) and Sierra et al. (2010) using the complete protein sequences from *P. veronii* (1YdBTEX2 and 1YB2), *P. fluorescens* (SBW25, A506, Pf0-1 and F113), *P. protegens* (CHA0 and Pf-5), *P. brassicacearum* NFM421 and *P. aeruginosa* PAO1 retrieved from the NCBI ftp website (http://www.ncbi.nlm.nih.gov/Ftp). Pairwise comparisons between *P. veronii* and *P. fluorescens* strains were done through an in-house perl script detecting the reciprocal best hit (RBH). The alignment within the RBH (pair of proteins) was scored using the ratio (Length of the alignment * %ID) / Length of the Query protein).

Two data matrices consisting of either a list of proteins or nucleotides from each of the 10 genomes (59,527 proteins) were generated and the identity/similarity score measured as the ratio (Length of the alignment * %ID) / Length of the Query protein) described above. The higher the ratio scores between a pair of genomes, the more related these genomes. A similarity matrix was generated using the Bray–Curtis coefficient (Bray & Curtis, 1957) for each data matrix and a respective dendogram generated by group-average agglomerative hierarchical clustering. All statistical routines were carried out using PRIMER 6 (v.6.1.6, PRIMER-E; Plymouth Marine Laboratory, UK) (Clarke & Warwick, 2001).

3.3.4 Partial purification of benzene dihydrodiol dehydrogenase

Cells of *P. veronii* 1YdBTEX2 and *P. veronii* 1YB2 (100 ml of culture) were harvested during late exponential growth, resuspended in 1 ml of 50 mM Tris-HCl buffer (pH 7.5) and disrupted with a French press (Aminco, Silver Spring, Md.). Cell debris were removed by ultracentrifugation at 100,000 × *g* at 4°C for 1 h. Benzene dihydrodiol dehydrogenase was partially purified by using a Fast Protein Liquid Chromatography system (Amersham Biosciences, Freiburg, Germany). Cell extracts containing 1µg of protein were applied directly onto the column, and proteins were eluted by using a linear gradient of 0 to 0.5 M NaCl, over 33 ml at a flow rate of 0.5 ml/min. Activity of benzene dihydrodiol dehydrogenase was determined spectrophotometrically by measuring the 3,5 cyclohexadiene 1,2 diol (benzene dihyrodiol 0.1 mM) dependent

reduction of NAD^+ (0.1 mM) in Tris-HCl (100 mM, pH 7.5) at 340 nm (ε_{NADH}= 6,220 $M^{-1} \cdot cm^{-1}$). Protein concentrations were determined by the method of Bradford (Bradford, 1976) with bovine serum albumin as the standard.

3.3.5 Identification of benzene dihydrodiol dehydrogenase

Aliquots of active fractions were subjected to SDS-PAGE on a Bio-Rad Miniprotein II as described previously (Laemmli, 1970). The acrylamide concentrations for concentrating and separating gels were 5 and 10% (wt/vol) respectively. The proteins were stained by Coomassie brilliant blue R250 and further electroblotted onto a polyvinylidene difluoride membrane, stained with Coomassie brilliant blue R250. Major protein bands with a molecular mass of ~50kDa were analyzed by N-terminal amino acid sequencing, performed with an Applied Biosystems model 494 A Procise HT sequencer as previously described (Heim et al., 2002). Proteins annotated by RAST from *P. veronii* 1YdBTEX2 and 1YB2 were searched for the presence of the determined N-terminal sequence.

3.3.6.Preparation of standards for real-time quantitative PCR (RT-q-PCR)

Primers were designed annealing inside genes of interest and targeting genes encoding the α-subunit of the isopropylbenzene dioxygenase (ipbAfw [5'-GCTGGTTCAATGATGACTTCTC-3'] and ipbArev [5'-GATGAAAGACCAGACTTCGAC-3']), benzene dihydrodiol dehydrogenase (DHfw [5'-GTGATTTTCTGCTCGAAGCTG-3'] and DHrev [5'-ATCCCAGGTCATCGCCGTAC-3']), extradiol dioxygenase EXDOA-1 (ExdoA-1fw [5'-GTTGACAAATTCTCCGTGGT-3'] and ExdoA-1rev [5'-TCTGCATACAACTCGAAGTGA-3']) and EXDO A-2 (ExdoA-2fw [5'-GCGTGTACTCGATATGGACA-3'] and ExdoA-2rev [5'-TGGGTCAGCTGATACAGACA -3']) (only in case of 1YdBTEX2). Gene fragments of interest were amplified using genomic DNA as template. PCR products were visualized on 1% agarose gels and purified using QIAEX II Gel Extraction KitQiagen kit (Qiagen, Germany). Excised fragments were cloned in the pGEM®-T Easy Vector (Promega, USA) following the manufacturer's procedure. Vectors with inserts were transformed into *E. coli* JM109, cells incubated over night and screened by PCR for the presence of the

inserts. Appropriate plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen, Germany).

In vitro RNA transcription was performed from plasmid DNA taking advantage of the T7 promoter sequence present in the pGEM-T vector through the addition of 10 µl of 10x transcription buffer, 10 µl NTPs (10 mM each) and 4 µl of T7 RNA polymerase (20 U/µl, Ambion) and incubation for 18 h at 37°C. After incubation with DNAse, cDNA standards of above mentioned gene transcripts were prepared through in-vitro retrotranscription using SuperScript III (Invitrogen) following the instructions of the manufacturer.

3.3.7 RNA extraction and reverse transcription

For RNA preparation, strain 1YB2 was grown with benzene or toluene, and 1YdBTEX2 with toluene as carbon sources (inducing conditions) and on glucose (20g/l) as control (non-inducing conditions). Three ml of culture were harvested during exponential growth and centrifuged at 4 °C and 3000 x g for 2 min. The pellet was resuspended in 1 ml of RNAproTM Solution (MP Biomedicals), placed into Lyzing Matrix B tubes (MP Biomedicals), lysed in a Fast Prep®-24 instrument (40 sec, intensity 6.0) and washed with an equal volume of chloroform. After centrifugation, 3 volumes of ethanol (-20 °C) were added to the aqueous phase and the solution transferred to an RNeasy Mini Kit column (Qiagen, Germany). The column was washed according to the manufacturer's instructions and RNA eluted in 60 µl RNase free water. Residual co-extracted genomic DNA was digested by incubation with turbo DNase (Ambion) following the manufacturer's instructions and the RNA purified using the RNeasy Mini Kit (Qiagen). The quality and quantity of RNA was analyzed spectrophotometrically by determination of the A260/A280 ratio. Subsequently, 800 ng of RNA was mixed with 1 µl of random primers (Invitrogen) and incubated for 10 min at 70°C. The reaction mixture was supplemented with 12 µl of 5x first strand buffer (Invitrogen), 6 µl 0.1 M DTT, 2 µl dNTP (10 mM), 8.8 µl of RNase free water and 2 µl of Superscript III RT (200 U/µl) and incubated for 2 h at 50°C, 30 min at 55°C and for 10 min at 70°C to inactivate the enzyme. The cDNA was purified with the PCR cleanup kit (Qiagen) and eluted in 100µl of DNAse free water.

3.3.8 Real Time Quantitative PCR (RT-q-PCR)

RT-q-PCR was used to analyze the transcript levels of isopropylbenzene dioxygenase, benzene dihydrodiol dehydrogenase and extradiol dioxygenase encoding genes involved in the degradation of benzene and toluene by 1YdBTEX2 and of benzene by 1YB2.

Primers (described above) targeting the α-subunit of isopropylbenzene dioxygenase, benzene dihydrodiol dehydrogenase and extradiol dioxygenases EXDO A-1 and A-2 were used for the amplification of cDNA. Each reaction was performed in duplicate in a final volume of 20 µL containing 2.5 µL of each primer (10 pmol), 10 µL QuantiTect SYBR Green PCR Master Mix (Qiagen) and 5 µL of DNA template (1/10 dilution of original extract). Amplification was conducted in a LightCycler® 480 Real-Time PCR System programmed to hold at 95 °C for 10 min, and to complete 50 cycles of 94 °C for 15 s, 56 °C for 30 s and 72 °C for 30 s. The PCR results were analyzed by the LightCycler® 480 Software (Roche Applied Science, USA). Standard curves were generated from serial dilutions of standards with known concentrations of cDNA by plotting threshold cycles (CT values) versus copy number. A melting step was added to confirm amplification specificity.

3.3.9 Accession number

The genomes sequences from *P. veronii* 1YdBTEX2 and 1YB2 have been deposited in GenBank under the accession numbers AOUH00000000 (Lima-Morales et al., 2013) and JGYI00000000, respectively.

3.4 Results and Discussion

3.4.1 General characteristics of *P. veronii* 1YdBTEX2 and 1YB2

P. veronii 1YdBTEX2 and 1YB2 genomes were sequenced and approximately 9.5 and 3.2 million reads were obtained, respectively. The sequences were assembled according to Lima-Morales et al. (2013), yielding 63 and 115 contigs for strains 1YdBTEX2 and 1YB2, respectively. The genomes contained 5,981 (1YdBTEX2) and 6,884

(1YB2) candidate protein-encoding genes, where more than 5,100 of the encoded proteins are shared between the strains (having a sequence identity ≥80%, at the nucleotide level). This corresponds to 85% and 74%, respectively of the proteins encoded in the genomes of 1YdBTEX2 and 1YB2. By functional comparison of genome sequences available, the RAST server (Aziz et al. 2008) revealed the closest neighbors of 1YdBTEX2 and 1YB2, with complete genome sequence, as *P. fluorescens* A506 (Loper et al., 2012) and *P. fluorescens* SBW25 (Silby et al., 2009). A comparison of finished genomes of strains of the *P. fluorescence* group (Table 1) shows that they vary in size from 5.51 –7.64 Mb, with protein coding genes ranging from 4796–6879.

Table 1. Genome information of *P. veronii* 1YdBTEX2, *P. veronii* 1YB2, *P. fluorescens* SBW25, *P. fluorescence* A506, *P. fluorescens* Pf0-1, *P. fluorescens* F113, *P. protegens* Pf-5, *P. protegens* CHA0, *P. brassicacearum* NFM1421 and *P. poae* RE*1-1-14.

	Features	Number of contigs	Genome size (Mb)	Plasmid size (Kb)	Gene count	Protein coding genes	GC content (%)	Genome sequence
P. veronii	1YdBTEX2	63	6.68	ND	6052	5981	59%	Lima-Morales et al. (2013)
	1YB2	115	7.64	ND	6943	6879	60%	this study
P. fluorescens	SBW25	1	7.14	42.5	6492	6395	60%	Silby et al. (2009)
	A506	1	5.96	5.69	5426	5323	60%	Loper et al. (2012)
	Pf0-1	1	6.44	-	5829	5722	61%	Silby et al. (2009)
	F113	1	6.85	-	5962	5862	61%	Redondo-Nieto et al. (2013)
P. protegens	Pf-5	1	7.07	-	6257	6142	63%	Paulsen et al. (2005)
	CHA0	1	6.86	-	6199	6116	63%	Jousset et al. (2014)
P. brassicacearum	NFM421	1	6.84	-	6176	6095	61%	Ortet et al. (2011)
P. poae	RE*1-1-14	1	5.51	6.37	4877	4796	61%	Müller et al. (2013)

ND, not determined

3.4.2 Phylogenetic analyses

Comparison of the 16S rRNA gene showed that the genes of strains 1YB2 and 1YdBTEX2 differed by only two bases over the 1,513 base stretch analyzed and both showed a 99.7% sequence identity with the type strain of *P. veronii* CIP104663^{T}. As only 98.7 of sequence identity was observed with the type strain of *P. fluorescens*

IAM12022^T, this indicates that both 1YB2 and 1YdBTEX2 belong to the *P. veronii* species rather than to the *P. fluorescens* species (see Figure 1A). However, 16S rDNA sequence identity between the different fully sequenced strains still termed *P. fluorescens* and *P. fluorescens* IAM12022^T varies between 96.6 and 98.8% indicating that at least some of them belong to different species. The gene encoding DNA gyrase subunit B (*gyrB*), is considered as a more selective phylogenetic marker (Yamamoto & Harayama, 1998) and a 798 base stretch of sequence is available for all fully sequenced *P. fluorescence* strains and the type strains of *P. fluorescens* and *P. veronii*. The high sequence identity of 99.2% to *P. veronii* CIP104663^T indicate that both 1YB2 and 1YdBTEX2 belong to the *P. veroniii* species (Figure 1B). In contrast, sequence identity to *P. fluorescens* IAM12022 and other microorganisms shown in Figure 1B is only 90-91%.

Among strains previously classified as *P. fluorescens*, strain Pf-5 was recently reclassified as *Pseudomonas protegens*, and accordingly, high sequence identity was observed in its gyrB sequence with *P. protegens* CHA01 (99.1 %). As indicated by analysis of the *gyr*B gene sequences, *P. fluorescens* A506 and SBW25 were grouped together in a multilocus sequence analysis (MLSA) whereas strain Pf0-1 clustered in another group (F113 was not analyzed in this study) (Loper et al., 2012), supporting the idea that *P. fluorescens* isolates belong to different species. This is in agreement with statements from genomic analyses, which already indicated strains Pf-5, Pf0-1 and SBW25 to belong to different species (Silby et al., 2009).

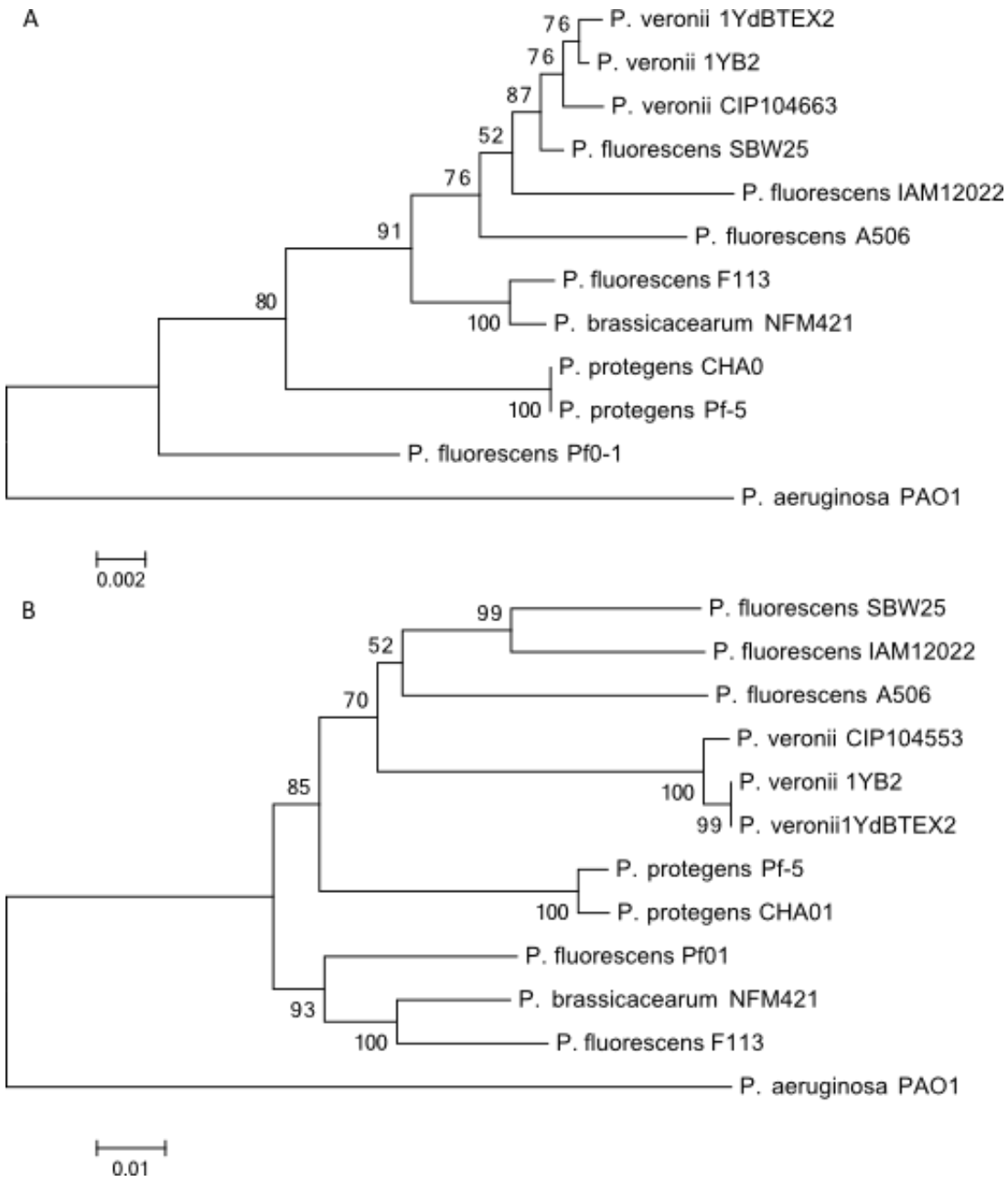


Figure 1: Phylogenetic relationships among *P. aeruginosa* PAO1, *P. veronii* 1YdBTEX2, 1YB2 and the *P. veronii* type strain CIP104663^{T}, *P.fluorescens* A506, SBW25, F113, Pf01 and the *P. fluorescens* type strain IAM12022^{T}, *P. brassicacearum* NFM421, *P. protegens* CHA01 and Pf-5based on the analysis of (A) 16S rRNA gene sequences and (B) DNA gyrase subunit B gene sequences (*gyrB*). Alignments were performed with MUSCLE 3.7 using default values. Phylogenetic trees were constructed with MEGA5 (Tamura et al) using the Neighbour-Joining (N-J) algorithm (Saitou & Nei, 1987) with Jukes-Cantor correction and pairwise deletion of gaps and missing data. A total of 100 bootstrap replications were done to test for branch robustness.

To analyze in more detail the phylogenetic relationship of 1YB2 and 1YdBTEX2 with previously genome sequenced isolates, the predicted proteins of these strains were compared with those of *P. fluorescens* (SBW25, A506, Pf0-1 and F113), *P.*

protegens (CHA0 and Pf-5), *P. brassicacearum* NFM421 and *P. aeruginosa* PAO1. Comparison of the similarity of total predicted proteins confirmed that the isolates under study here are more closely related to *P. fluorescence* A506 and SBW25 (Figure 2) than to *P. fluorescens* F113 or Pf01. Also, the analysis clearly shows the high similarity between both *P. protegens* strains CHA0 and Pf-5, while *P. fluorescens* F113 and *P. brassicacearum* NFM421 cluster together, which may further indicate that, as already seen from comparisons of 16S rDNA and gyrase B encoding gene sequences, that both strains belong to the same species. The high similarity of the *P. veronii* isolates is somehow masked by the fact that especially strain 1YB2 encodes various proteins not observed in 1YdBTEX2. If these are encoded on mobile genetic elements remains to be elucidated.

Pairwise comparisons of proteins encoded by these strains are shown in figure 3 (excluding *P. aeruginosa* PAO1) and support the assumption that strains 1YB2 and 1YdBTEX2, as well as *P. protegens* CHA01 and Pf-5, as well as *P. fluorescens* F113 and *P. brassicacearum* belong to the same species, as the majority of proteins shared by these strains have a similarity higher than 90%.

Lastly, these robust genomic comparisons (at both the nucleotide and amino acid level) also confirm that strains previously identified as *P. fluorescens* can be very different to each other, suggesting that some of these strains should be reclassified, allowing a better organization of species representing the *Pseudomonas* genus.

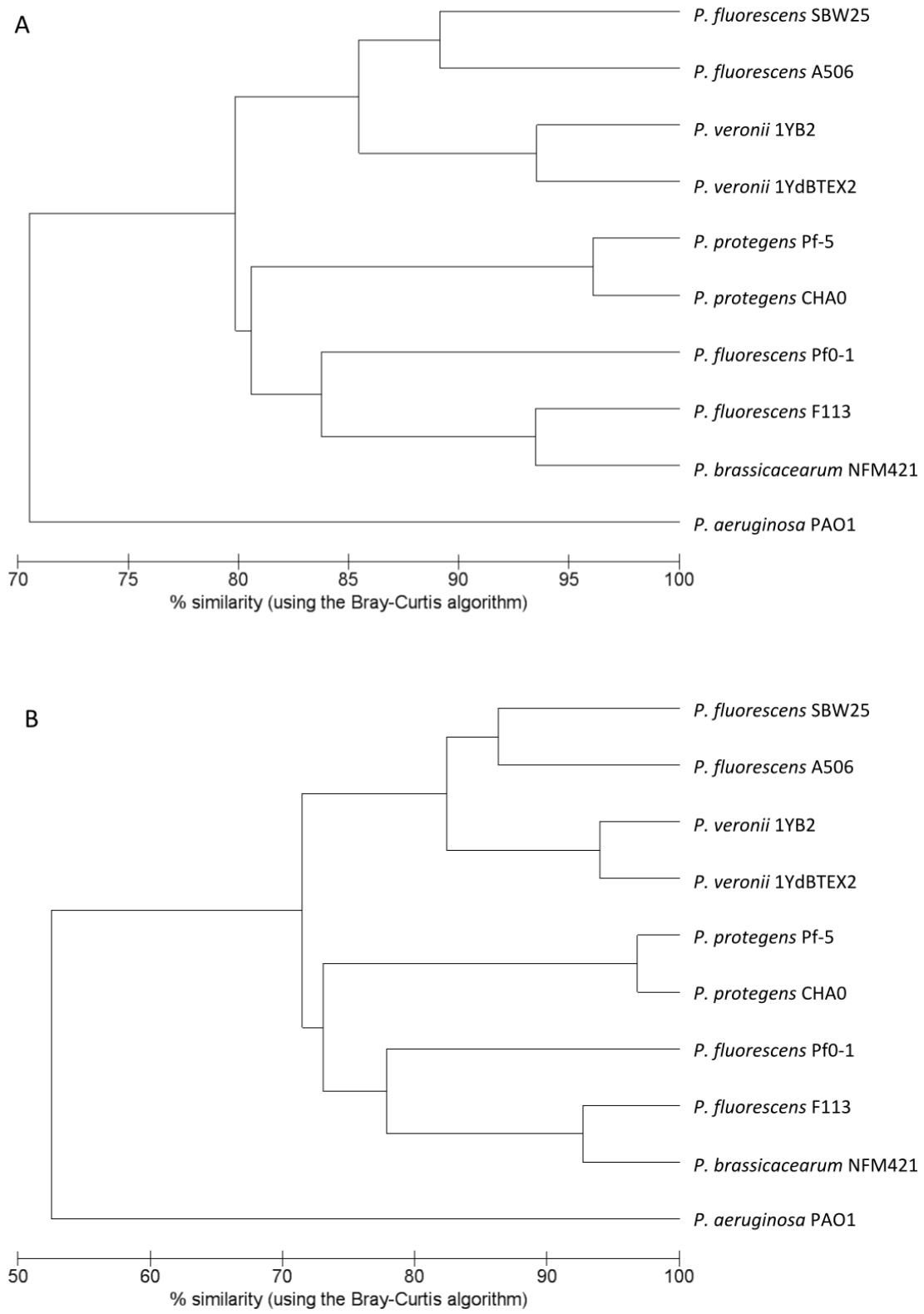
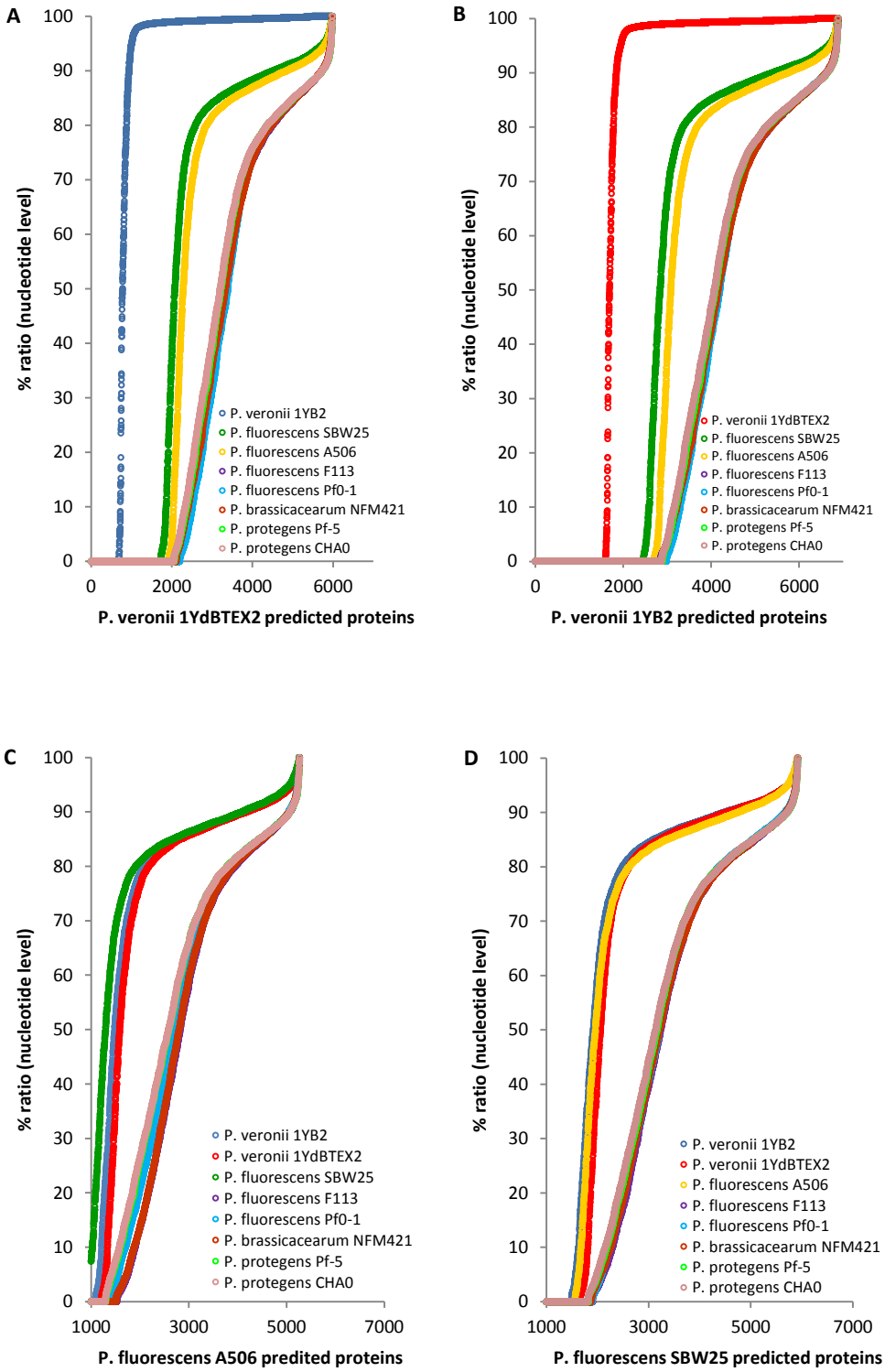
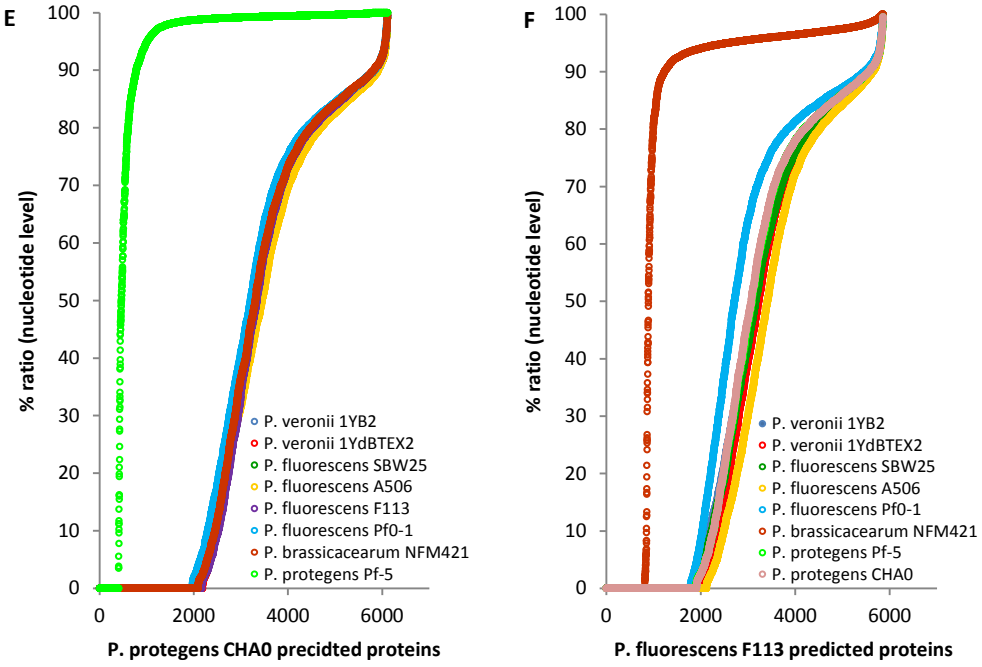
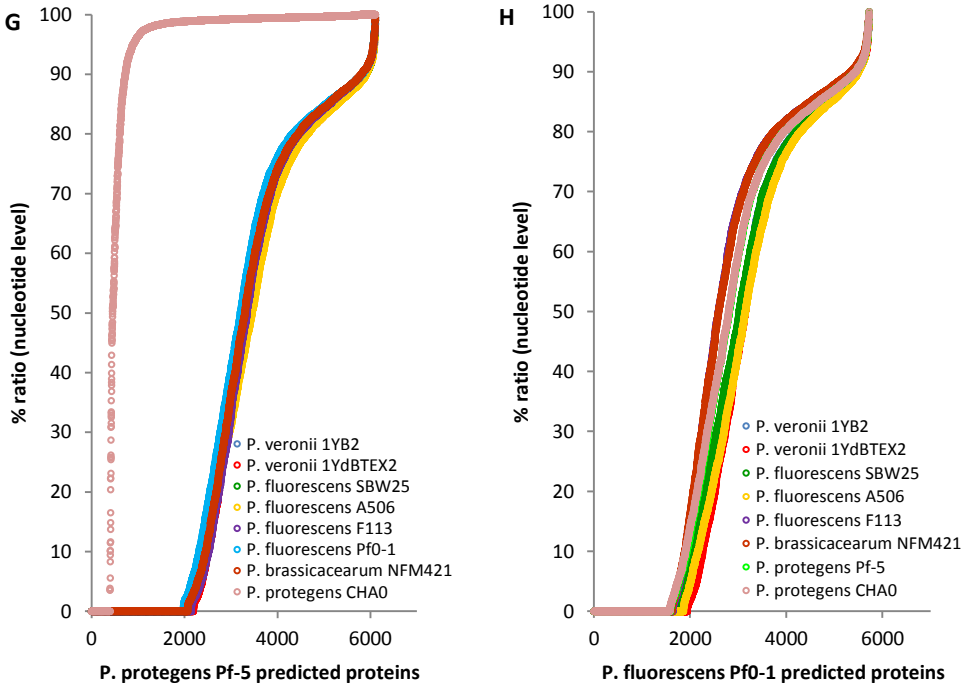


Figure 2: Genomic comparisons of *P. veronii* strain 1YdBTEX2 and 1YB2 with *P. fluorescens* SBW25, A506, F113 and Pf0-1, *P. brassicacearum* NFM421, *P. protegens* CHA0 and Pf-5 and *P. aeruginosa* PAO1 at the amino acid level (A) and at the nucleotide level (B). The cluster was determined by group-average agglomerative hierarchical clustering after applying the Bray-Curtis similarity algorithm.


A
% ratio (nucleotide level)
P. veronii 1YdBTEX2 predicted proteins
P. veronii 1YB2
P. fluorescens SBW25
P. fluorescens A506
P. fluorescens F113
P. fluorescens Pf0-1
P. brassicacearum NFM421
P. protegens Pf-5
P. protegens CHA0
B
% ratio (nucleotide level)
P. veronii 1YB2 predicted proteins
P. veronii 1YdBTEX2
P. fluorescens SBW25
P. fluorescens A506
P. fluorescens F113
P. fluorescens Pf0-1
P. brassicacearum NFM421
P. protegens Pf-5
P. protegens CHA0
C
% ratio (nucleotide level)
P. fluorescens A506 predited proteins
P. veronii 1YB2
P. veronii 1YdBTEX2
P. fluorescens SBW25
P. fluorescens F113
P. fluorescens Pf0-1
P. brassicacearum NFM421
P. protegens Pf-5
P. protegens CHA0
D
% ratio (nucleotide level)
P. fluorescens SBW25 predicted proteins
P. veronii 1YB2
P. veronii 1YdBTEX2
P. fluorescens A506
P. fluorescens F113
P. fluorescens Pf0-1
P. brassicacearum NFM421
P. protegens Pf-5
P. protegens CHA0


E
% ratio (nucleotide level)
P. protegens CHA0 precidted proteins
P. veronii 1YB2
P. veronii 1YdBTEX2
P. fluorescens SBW25
P. fluorescens A506
P. fluorescens F113
P. fluorescens Pf0-1
P. brassicacearum NFM421
P. protegens Pf-5
F
% ratio (nucleotide level)
P. fluorescens F113 predicted proteins
P. veronii 1YB2
P. veronii 1YdBTEX2
P. fluorescens SBW25
P. fluorescens A506
P. fluorescens Pf0-1
P. brassicacearum NFM421
P. protegens Pf-5
P. protegens CHA0


G
% ratio (nucleotide level)
P. protegens Pf-5 predicted proteins
P. veronii 1YB2
P. veronii 1YdBTEX2
P. fluorescens SBW25
P. fluorescens A506
P. fluorescens F113
P. fluorescens Pf0-1
P. brassicacearum NFM421
P. protegens CHA0
H
% ratio (nucleotide level)
P. fluorescens Pf0-1 predicted proteins
P. veronii 1YB2
P. veronii 1YdBTEX2
P. fluorescens SBW25
P. fluorescens A506
P. fluorescens F113
P. brassicacearum NFM421
P. protegens Pf-5
P. protegens CHA0

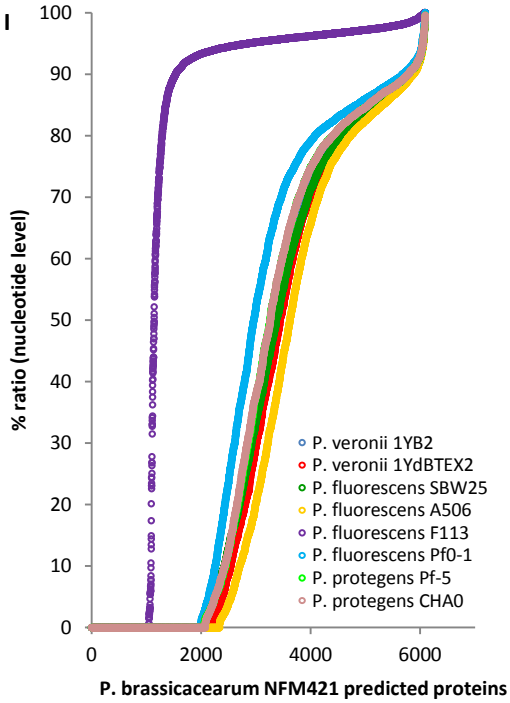


Figure 3: Comparison of proteins encoded by the genomes of *P. veronii* 1YdBTEX2 and 1YB2, *P. fluorescens* SBW25, A506, F113 and Pf0-1, *P. brassicacearum* NFM421, *P. protegens* CHA0 and Pf-5 were performed using the following genomes as reference: (A) *P. veronii* 1YdBTEX2, (B) *P. veronii* 1YB2, (C) *P. fluorescence* SBW25, (D) *P. fluorescence* A506, (E) *P. protegens* CHA0, (F) *P. fluorescence* F113, (G) *P. protegens* Pf-5, (H) *P. fluorescence* Pf0-1 and (I) *P. brassicacearum* NFM421. Each circle represents one predicted protein, shared between the reference strain and the strains being compared. The y-axis shows the percentage of homology (at the nucleotide level) of each predicted protein found in both, the reference strains and the strains that are been compared. The X-axis represents cumulative number of predicted proteins found in the reference strain genome. Note that the closer the strain line to the y-axis, the more closely related the respective proteome is to that of the reference strains.

3.4.3 The catabolic potential of *P. veronii* 1YdBTEX2 and 1YB2

Different catabolic gene clusters for the degradation of aromatic compounds were localized in both *P. veronii* 1YdBTEX2 and 1YB2 genomes (Figure 4).

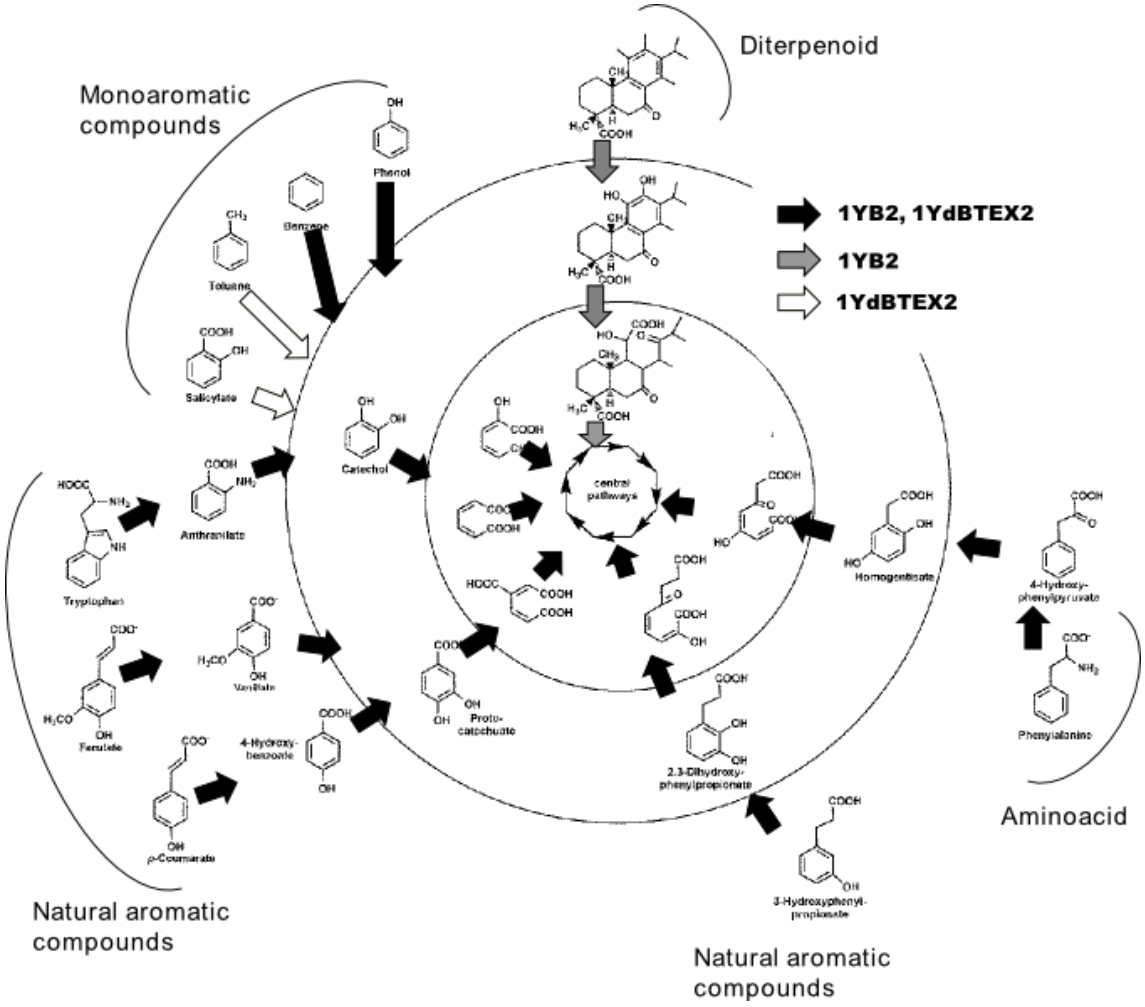


Figure 4: Overview of the degradation potential for aromatic compounds of *P. veronii* 1YdBTEX2 and 1YB2. Aromatic compounds are funneled through peripheral reactions into central intermediates, which are then processed by central pathways to TCA cycle intermediates. Black arrows indicate predicted pathways shared by *P. veronii* 1YdBTEX2 and 1YB2, white arrows indicate predicted pathways only found in 1YdBTEX2 and grey arrows indicate predicted pathways only found in 1YB2.

Both strains harbor complete phenol catabolic gene clusters comprising a multicomponent phenol monooxygenase and both can grow on this compound as sole carbon and energy source. The cluster shows an organization identical to that observed in *P. putida* MT4 (Takeo et al., 2006). Only strain *P. veronii* 1YdBTEX2 shows the presence of a salicylate catabolic gene cluster comprising a flavoprotein salicylate

1-hydroxylase and accordingly, only 1YdBTEX2 can grow with this compound. None of the strains was capable to grow on naphthalene, in accordance with the absence of a respective gene cluster in 1YdBTEX2 and the presence of only a truncated gene cluster in 1YB2 (Figure 5). Diterpenoid degradation has been well studied in *Pseudomonas abietaniphila* BKME-9 (Martin & Mohn, 1999, 2000) and *Burkholderia xenovorans* LB400 (Smith et al., 2007). A diterpenoid catabolic gene cluster is only present in 1YB2. The BKME-9 diterpenoid degradation cluster contains 20 genes (Smith et al., 2004). The *ditA1A2A3* encode the α- and β-subunits and the ferredoxin of the dioxygenase which transforms 7-oxodehydroabietic acid into 7-oxo-11,12-dihydroxy-8,13-abietadien acid. Other genes of this cluster where the gene products have been analyzed in detail comprise *ditC*, encoding an extradiol dioxygenase cleaving 7-oxo-11,12-dihydroxydehydroabietic acid and *ditQ*, encoding a cytochrome P450 monooxygenase probably transforming dehydroabietic acid into 7-hydroxy dehydroabietic acid (Martin & Mohn, 2000; Smith et al., 2004, 2008). The cluster found in 1YB2 has an organization identical to as that observed in BKME-9, except for the absence of *ditP*, a hypothetical protein and the presence of *ditU*, a second cytochrome P450 which is important for abietic acid, but not dehydroabietic acid and which was only found in the operon of LB400 (Smith et al., 2008). The aminoacid sequence identity between the 1YB2 *ditA1A2A3BC* encoded proteins and those encoded by the *P. abietaniphila* BKME-9 genome (88-97%) is significantly higher than the sequence identity with the *B. xenovorans* LB400 encoded proteins (51-75%). Interestingly 1YB2 harbours a second cluster comprising only *ditA1A2C* genes, which is located approximately 57 kb apart from the complete *dit* gene cluster, where the encoded proteins exhibit 52-72% of sequence identity to proteins of the complete *dit* gene cluster. Actually *ditA1* from the the incomplete cluster found in strain 1YB2 was described before (Witzig et al., 2007).

Genetic analysis of determinants of the benzene catabolic pathway

During previous studies, part of the isopropylbenzene dioxygenase encoding genes (*ipb*) (Witzig et al., 2006) and the genes encoding an extradiol dioxygenase with high similarity to that of the archaetype TOL plasmid encoded catechol *meta*-cleavage

pathway (termed EXDO A) involved in the degradation of benzene by both strains had already been identified (Junca et al., 2004). Genome sequencing indicated the presence of a complete isopropylbenzene catabolic operon in both strains (Figure 5), including a gene encoding an isopropylcatechol 2,3-dioxygenase (termed EXDO K2 thereafter) and a gene organization similar to the one found on plasmid pRE4 of *P. putida* RE204 (Eaton et al, 1998). The presence of transposases flanking the *ipb* operons in both 1YB2 and 1YdBTEX2 indicate that they were acquired via horizontal gene transfer. Curiously, activity of the encoded extradiol dioxygenase has never been observed (Junca et al., 2004). A detailed analysis revealed a point mutation in the extradiol dioxygenase encoding gene of 1YB2, resulting in an internal stop codon, and thus a non-functional gene product. In 1YdBTEX2, the extradiol dioxygenase encoding gene was obviously functional, however further sequence analysis showed point mutations in the isopropylbenzene dihydrodiol dehydrogenase (*ipbB*) encoding gene upstream of the extradiol dioxygenase encoding gene, resulting in stop codons in both strains.

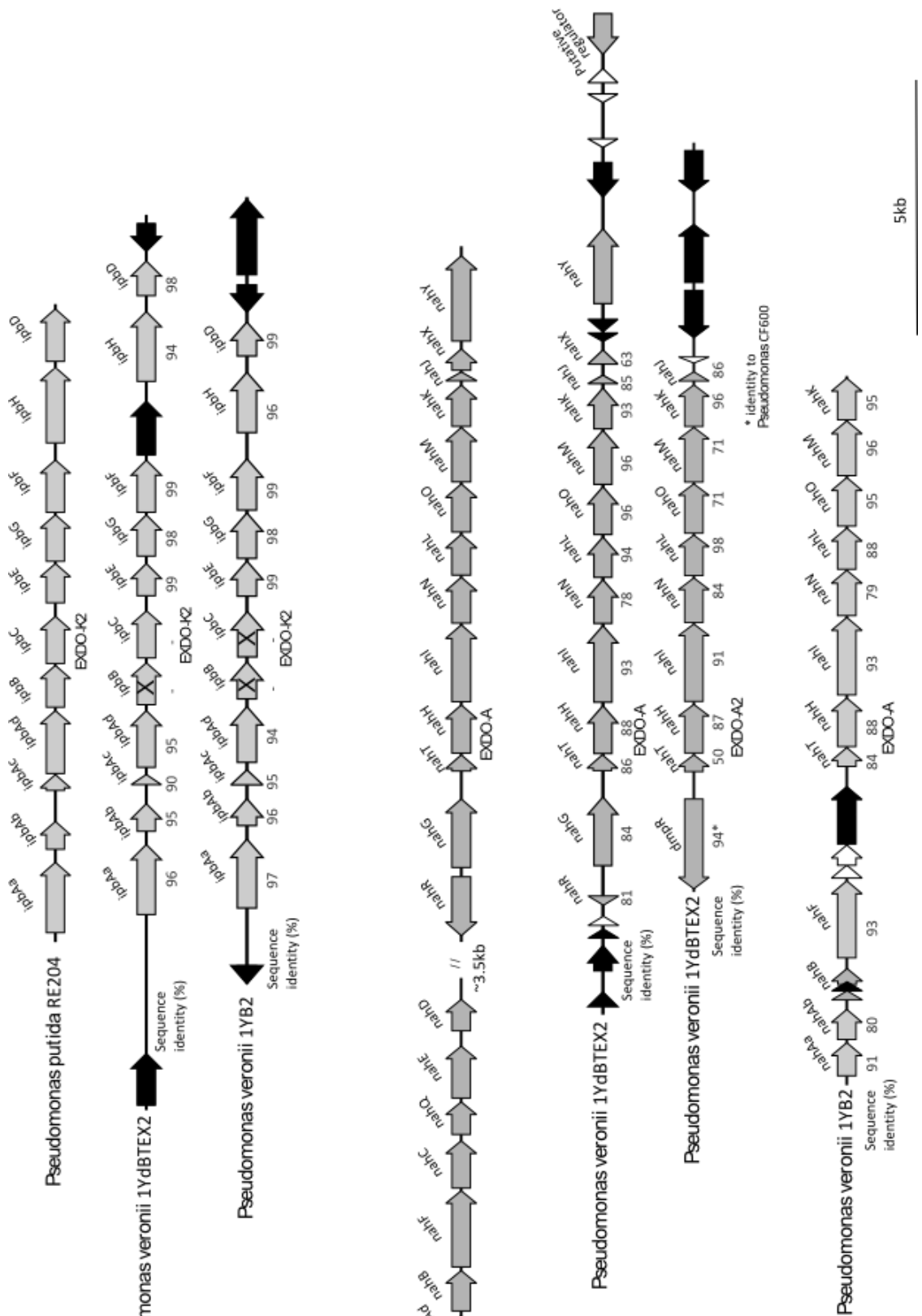

(A)
Pseudomonas putida RE204
EXDO-K2
Pseudomonas veronii 1YdBTEX2
Sequence identity (%)
EXDO-K2
Pseudomonas veronii 1YB2
Sequence identity (%)
EXDO-K2
(B)
Pseudomonas putida G7 plasmid Nah7
~3.5kb
EXDO-A
Pseudomonas veronii 1YdBTEX2
Sequence identity (%)
EXDO-A
Putative regulator
Pseudomonas veronii 1YdBTEX2
Sequence identity (%)
EXDO-A2
* identity to Pseudomonas CF600
Pseudomonas veronii 1YB2
Sequence identity (%)
EXDO-A
5kb

Figure 5: Organization of the clusters encoding genes involved in benzene degradation by *P. veronii* 1YdBTEX2 and 1YB2. Gene clusters of *P. putida* RE204 and *P. putida* G7 were used for comparison. (A) Isopropylbenzene gene cluster (*ipb*) of *P. putida* RE204 (Eaton et al, 1998); *ipb* gene cluster of *P. veronii* 1YdBTEX2 where a cross indicates a stop codon in the *ipbB* dihydrodiol dehydrogenase encoding gene; *ipb* gene cluster of *P. veronii* 1YB2, where crosses indicate stop codons in the *ipbB* dihydrodiol dehydrogenase and EXDO K2 extradiol dioxygenase encoding genes; (B) Upper and lower naphthalene catabolic gene clusters (*nah*) located on plasmid Nah7 of *P. putida* G7 (Yen and Gunsalus, 1982); gene cluster in *P. veronii* 1YdBTEX2 encoding enzymes for salicylate degradation; gene cluster in *P. veronii* 1YdBTEX2, harboring a second EXDO A (EXDO A2) encoding gene; truncated *nah* gene cluster in *P. veronii* 1YB2. Black and white arrows indicate mobile genetic elements and hypothetical proteins, respectively. The percent sequence identity between the encoded proteins of *P. putida* RE204 and *P. veronii* 1YdBTEX2 or 1YB2 (A) and between the encoded proteins of *P. putida* G7 and *P. putida* 1YdBTEX2 or 1YB2 (B) are given below the gene clusters.

The third step in benzene degradation is the cleavage of catechol, for which, as described before, an EXDO A is recruited in *P. veronii* 1YB2 and 1YdBTEX2 (Junca et al., 2004). Sequence comparisons show that in *P. veronii* 1YdBTEX2 EXDO A is encoded by a salicylate catabolic gene cluster, which is recruited for benzene degradation. The presence of a complete salicylate catabolic gene cluster allows the strain to also grow on this substrate, as described above. Typically, salicylate catabolic gene clusters are localized in the proximity of naphthalene catabolic gene clusters, as is the case in *P. putida* G7 (Yen and Gunsalus, 1982). In this strain two catabolic gene operons are localized on plasmid NAH7, where the first operon includes the genes *nahAaAbAcAdBCDEF*, encoding enzymes for the transformation of naphthalene to salicylate and the second cluster comprising *nahGTHINLOMKJ* genes responsible for salicylate oxidation via the catechol *meta*-cleavage (EXDO-A) pathway (Figure 5). However, there was no indication for the presence of a naphthalene operon in *P. veronii* 1YdBTEX2. In contrast, in *P. veronii* 1YB2 the EXDO-A encoding gene was localized in an incomplete naphthalene gene cluster (Figure 5). Exploration of this gene cluster revealed that the typical *nahAaAbAcAdBCDEF* structure was truncated. The naphthalene dihydrodiol dehydrogenase encoding gene (*nahB*) is interrupted by a transposase and the genes encoding the initial multicomponent naphthalene dioxygenase are incomplete, which explains the failure of the strain to grow on naphthalene. Finally, the *nah* operon is composed only by *nahAaAbBF* genes and downstream an incomplete salicylate operon is found, where the *nahG* gene encoding

salicylate hydroxylase is absent, which explains the failure of the strain to grow on salicylate.

P. veronii 1YdBTEX2 harbors a second extradiol dioxygenase (EXDO A-2), which exhibits 84% of sequence identity with EXDO A-1, and which is localized in a gene cluster comprising only genes for downstream processing of catechol ring-cleavage products (Figure 5). The absence of any gene encoding enzymes for the formation of catechol may indicate the function of this operon in the processing of catechol derived from different sources by gene products encoded elsewhere on the genome.

Importantly, no functional dihydrodiol dehydrogenase was, thus, observed in any of the aforementioned gene clusters, and screening of the genome did not reveal any open reading frame which could encode for an enzyme with high similarity to those transforming isopropylbenzene dihydrodiol, naphthalene dihydrodiol or benzene dihydrodiol in either strain. We therefore decided to purify the dihydrodiol dehydrogenase that has been recruited for benzene degradation and to perform N-terminal sequencing to search deeper into the genome of both strains.

Identification of the benzene dihydrodiol dehydrogenase recruited for benzene degradation

The benzene dihydrodiol dehydrogenases of *P. veronii* 1YdBTEX2 and 1YB2 were partially purified from benzene grown cells and proteins in the active fractions were subjected to N-terminal sequencing. Only one open reading frame (ORF) of each genome matched the observed MQRFQNKVVLVTGAA N-terminal sequence. Exploring the gene organization, the benzene dihydrodiol dioxygenase encoding gene in both strains is localized in identical gene clusters, devoid of other genes encoding aromatic degradation enzymes (Figure 6). Importantly, the benzene dihydrodiol dehydrogenases of both 1YdBTEX2 and 1YB2 are only distantly related to previously described dehydrogenases involved in aromatic degradation pathways (Figure 6).

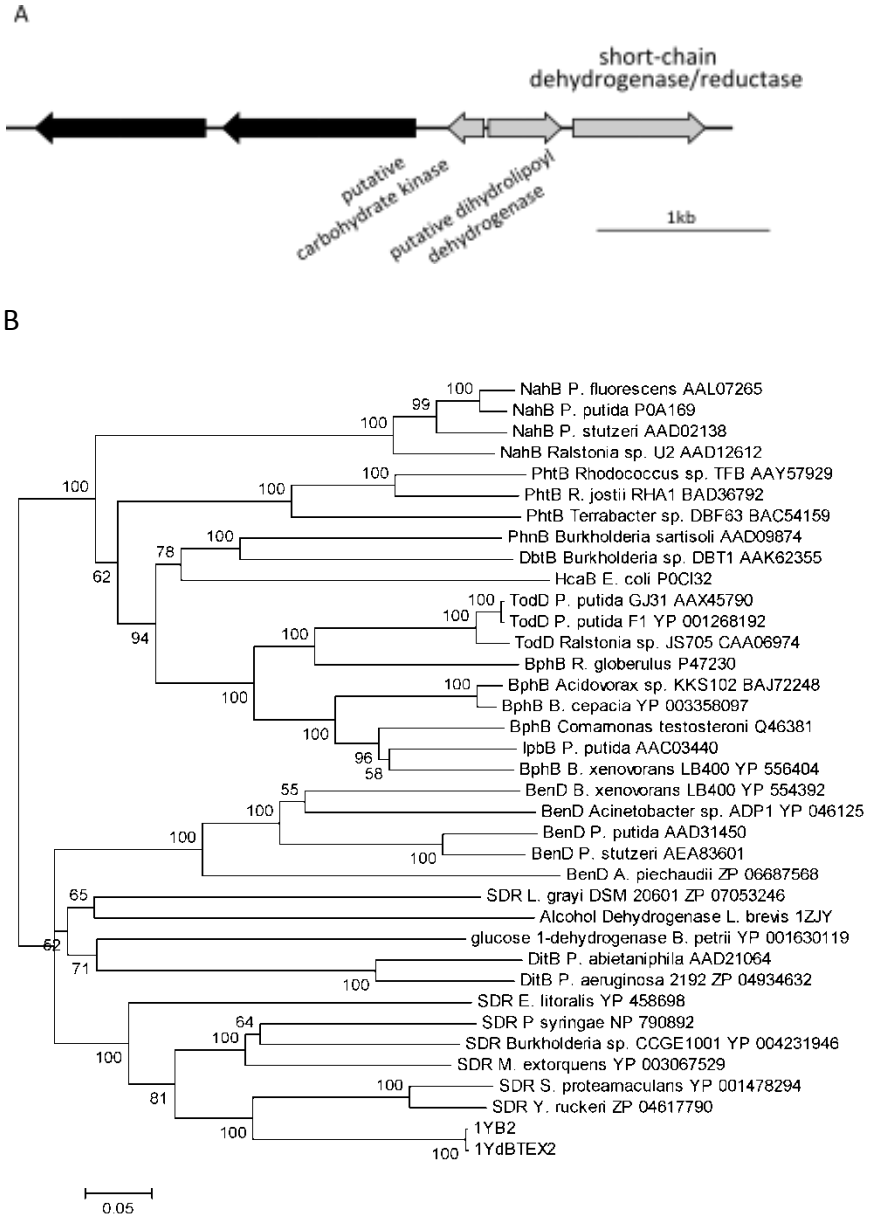


Figure 6: Identification of the benzene dihydrodiol dehydrogenase recruited for benzene degradation. (a) Gene cluster in *P. veronii* 1YdBTEX2 and 1YB2, which comprises a gene encoding a short-chain dehydrogenases/reductase recruited for benzene and/or toluene degradation. The black arrows represent mobile elements. (b) Evolutionary relationships of members of the short-chain dehydrogenases/reductase family of enzymes. The evolutionary histories were inferred using the neighbour-joining method and the p-distance model. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. Phylogenetic analyses were conducted in MEGA5 (Tamura et al., 2011). Bootstrap values above 50% from 100 neighbour-joining trees are indicated to the left of the nodes. Bar, 0.05 amino acid differences per site.

Expression of genes of the benzene catabolic pathway in P. veronii 1YdBTEX2 and 1YB2

According to the data presented above, benzene degradation in both strains is most likely catalyzed by proteins of at least three catabolic gene clusters. In addition, 1YdBTEX2 harbors two distinct extradiol dioxygenases of the EXDO A subfamily (EXDO A-1 and EXDO A-2). To analyze whether both dioxygenases were recruited and to determine whether all three gene clusters described above were transcribed in response to benzene and/or toluene, transcripts levels during growth on benzene (*P. veronii* 1YB2) and on benzene and toluene (*P. veronii* 1YdBTEX2) were compared to transcript levels observed under non-inducing conditions (growth on glucose).

Benzene dihydrodiol dehydrogenase and EXDO A were expressed constitutively in both strains with transcript levels of $2x10^5$ and $4x10^3$ copies/ng cDNA respectively. In contrast, IpbA and EXDO A2 (only present in 1YdBTEX2) were inducible with transcript levels being increased by 2 orders of magnitude after growth on aromatic substrates. However, even uninduced cells revealed high levels of transcripts of $4X10^2$ - $8X10^2$ copies/ng cDNA (Figure 7).

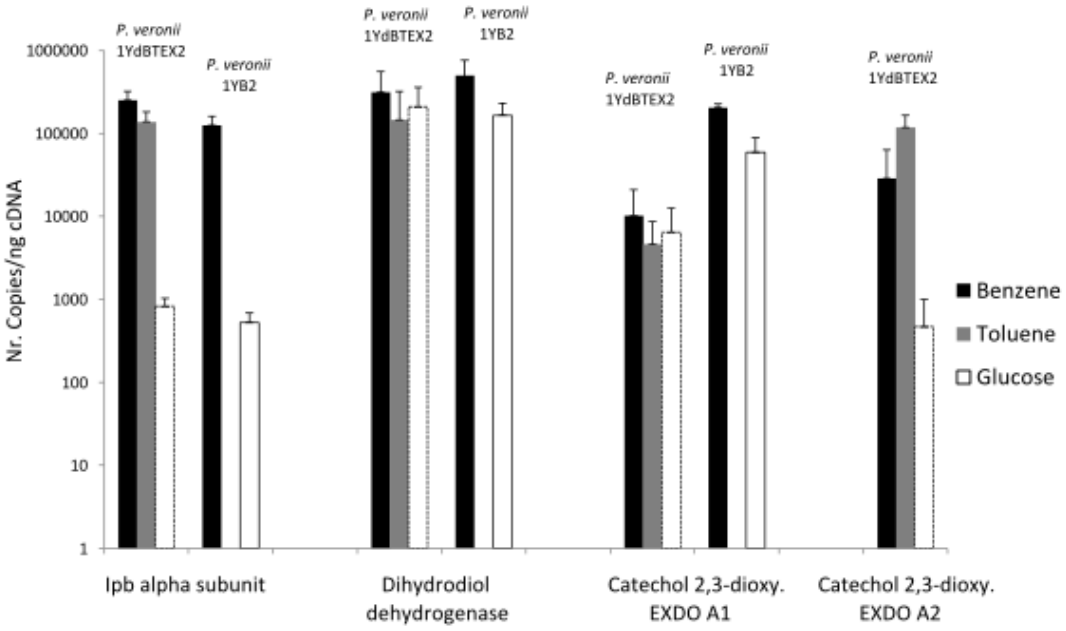


Figure 7: Accumulation of transcripts of catabolic genes measured during growth on benzene for both *P. veronii* 1YdBTEX2 and 1YB2 and on toluene only for *P. veronii* 1YdBTEX2. Glucose was used as a noninducing negative control.

3.4.4 Genetic attributes that may help *P. veronii* 1YdBTEX2 and 1YB2 to survive under environmental conditions

Siderophore production gene cluster

Pyoverdin is a siderophore excreted to aquire iron by all fluorescent pseudomonads, including *P. veronii* (Elomari et al, 1996) and respective biosynthetic gene clusters were also found in *P. veronii* 1YdBTEX2 and 1YB2. Pyoverdin has a high affinity to iron, however the number of proteins necessary for its biosyntheses is huge (Cornelis 2010). Pyochelin, a secondary siderophore with lower affinity to iron and a simpler structure was first isolated from *P. aeruginosa* PAO1 (ATCC 15692). An enantiomer of pyochelin (enantio-pyochelin) is synthesized by *P. protegens* CHA0 (Youard et al., 2007; Kupferschmied et al., 2013).

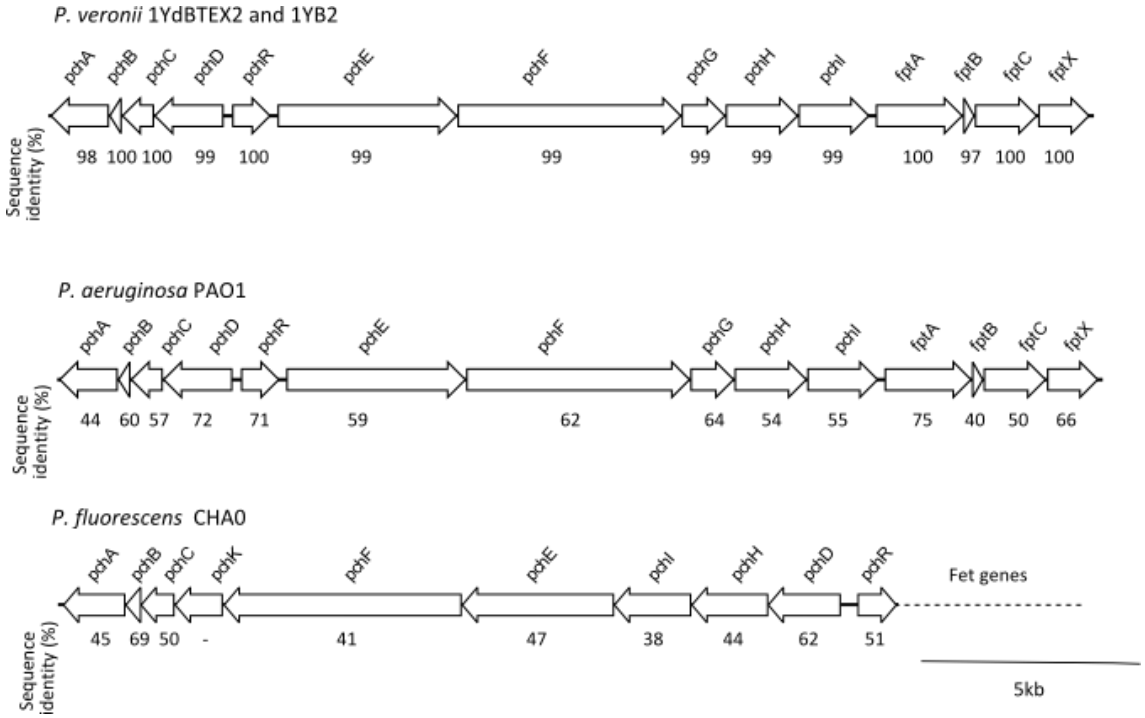


Figure 10: *P. veronii* 1YdBTEX2, 1YB2, *P. aeruginosa* PAO1 and *P. protegens* CHA0 pyochelin and enantiopyochelin biosynthetic gene clusters. The percent sequence identity between the encoded proteins of the *P. veronii* 1YdBTEX2 and 1YB2 gene clusters (top) as well as those between the encoded proteins of *P. veronii* 1YdBTEX2 on the one hand and *P. aeruginosa* PAO1 (middle) or *P. protegens* CHA0 (bottom) on the other are given below the gene clusters.

Pyochelin and enantiopyochelin are condensation products of salicylate and two cysteinyl residues and their synthesis require proteins encoded by two divergent operons *pchDCBA* and *pchEFGHI* (Serino et al., 1997; Reimmann et al., 1998). The PchA and PchB enzymes convert chorismate into salicylate (Gaille et al., 2002; Gaille et al., 2003), which is further coupled to the cysteine moieties by a thiotemplate mechanism involving the following enzymes: PchD, a salicylate adenylating enzyme, PchE and PchF two peptide synthetases and PchG, a reductase (Quadri et al., 1999; Patel & Walsh 2001; Reimmann et al., 2001). The transcription is regulated by the transcriptional regulator encoded by *pchR* (Patel & Walsh 2001; Crosa & Walsh 2002). Transport of the iron loaded forms of pyochelin or enantio-pyochelin is performed either by a single-subunit siderophore transporter encoded by *fptX*, as observed in PAO1 or a classical periplasmic binding protein-dependent ABC transporter (*fetCDE*) as observed in *P. protegens* CHA0 (Reimmann et al., 2012). Interestingly both *P. veronii* 1YdBTEX2 and 1YB2 harbor pyochelin gene clusters with the same organization as the one found in *P. aeruginosa* PAO1 with amino acid sequence identities of 40 % to 75% (Figure 8). Both *P. veronii* strains sequenced here harbour an *fpt* operon including an *fptX* transporter as described for PAO1 (Reimmann et al., 2012) (Figure 8).

The ability to produce the high affinity siderophore pyoverdine in addition to a second siderophore of lower affinity but simple synthesis may support *P. veronii* 1YB2 and 1YdBTEX2 in their competition with other micro-organisms and to survive in harsh environments.

A complete denitrification pathway is localized in one superoperonic gene cluster

Microorganisms that can use alternative electron acceptor for growth have advantage in anoxic environments. Denitrification, or the reduction of nitrate into dinitrogen gas, is a widespread capability among microorganisms belonging to phylogenetically distinct groups of bacteria and archaea. Bacteria capable of denitrification are frequently isolated from soil, sediment and aquatic environments (Philippot, et al., 2001). The complete denitrification pathway consists of the subsequent use of nitrate,

nitrite, nitric oxide and nitrous oxide (Philippot 2002) as terminal electron acceptors. Denitrification has being described in *Pseudomonas* species, e.g. from the 44 completely sequenced *Pseudomonas* spp in the Pseudomonas Genome Database (http://www.pseudomonas.com), 17 harbour the complete denitrification pathway among them all 9 *P. aeruginosa* and all 6 *P. stutzeri*, strains. In addition a complete denitrification pathway was observed in P*. brassicacearum* NFM421 and *P. fluorescence* F113, which as described above, may belong to the same species. Importantly, both *Pseudomonas veronii* 1YdBTEX2 and 1YB2 have all the factors necessary to carry out complete denitrification from nitrate to N_2 (Figure 9).

The nitrate reductase (Nar) gene clusters of *P. veronii* 1YdBTEX2 and 1YB2 show the same organization as observed in the well characterized *P. aeruginosa* PAO1 (Stover et al., 2000), *P. fluorescens* F113 (Redondo-Nieto et al., 2011) and *P. extremaustralis* 14-3 (Tribelli et al., 2012), where the highest amino acid sequence identity (>94%) was always observed with proteins originating from last mentioned strain. However, despite the high sequence similarity of proteins encoded by the *nar* cluster of *P. veronii* 1YB2, 1YdBTEX2 and *P. extremaustralis* 14-3 a *nir* (nitrite reductase) gene cluster is absent from last mentioned strain.

In *P. veronii* 1YB2 and 1YdBTEX2 the reduction of nitrite to nitric oxide is catalyzed by a cytochrome cd_1-nitrite reductase (*nirS)*, where the encoding gene is followed by *nirMCFDLGHJEN* genes, encoding enzymes involved in the maturation of NirS. The gene organization is identical to that of *P. aeruginosa* PAO1 (Figure 9) and *P. fluorescens* F113, with highest amino acid sequence identity of 68-87% being observed with last mentioned strain.

The nitric oxide reductase gene cluster (*nor* cluster) consists of *norC* and *norB* encoding the two structural subunits and other auxiliary genes. In *P. veronii* 1YdBTEX2 and 1YB2 a *norDBCEQ* gene cluster is observed in an organization similar to that in PAO1 and 14-3, however with an insertion between *norE* and *norC* of two genes. The highest amino acid sequence identity of the structural subunits is observed with *P. extremaustralis* 14-3 encoded proteins (e.g. 89% with NorC and 90% with NorB) (Figure 9).

The multicopper enzyme nitrous oxide reductase catalyses the reduction of nitrous oxide into dinitrogen gas. The *nos* cluster comprising genes encoding the structural gene (*nosZ)*, a regulatory protein (*nosR*) and accessory proteins required in the biosynthesis of the catalytically active enzyme (*nosDFYL)* shows an organization identical to that in *P. aeruginosa* PAO1, *P. fluorescens* F113 and *P. extremaustralis* 14-3 with highest identity of the encoded proteins of 87-96% to those encoded by *P. extremaustralis* 14-3.

For growth under oxygen-free atmosphere, some strict or facultative anaerobes such as *Pseudomonas* strains depend on a class III ribonucleotide reductase for the synthesis of deoxyribonucleotides (Siedow et al., 1999). In *P. veronii* 1YdBTEX2 and 1YB2 the *nrd* gene cluster encoding the glycyl-radical enzyme NrdD and the activating enzyme NrdG is located upstream of the *nar* operon with the encoded proteins again showing highest similarity *(*94-98%) with proteins encoded by *P. extremaustralis* 14-3.

A prominent difference between the *P. veronii* 1YdBTEX2 and 1YB2 denitrification cluster and that of other gene clusters encoding denitrification in *Pseudomonas* is that all 5 gene clusters crucial for complete denitrification of nitrogen gas (*nrd, nar, nir, nor* and *nos*) are organized in one superoperonic gene cluster. Figure 9, shows the distance in base pair between these gene clusters in PAO1, F113 and 14-3.

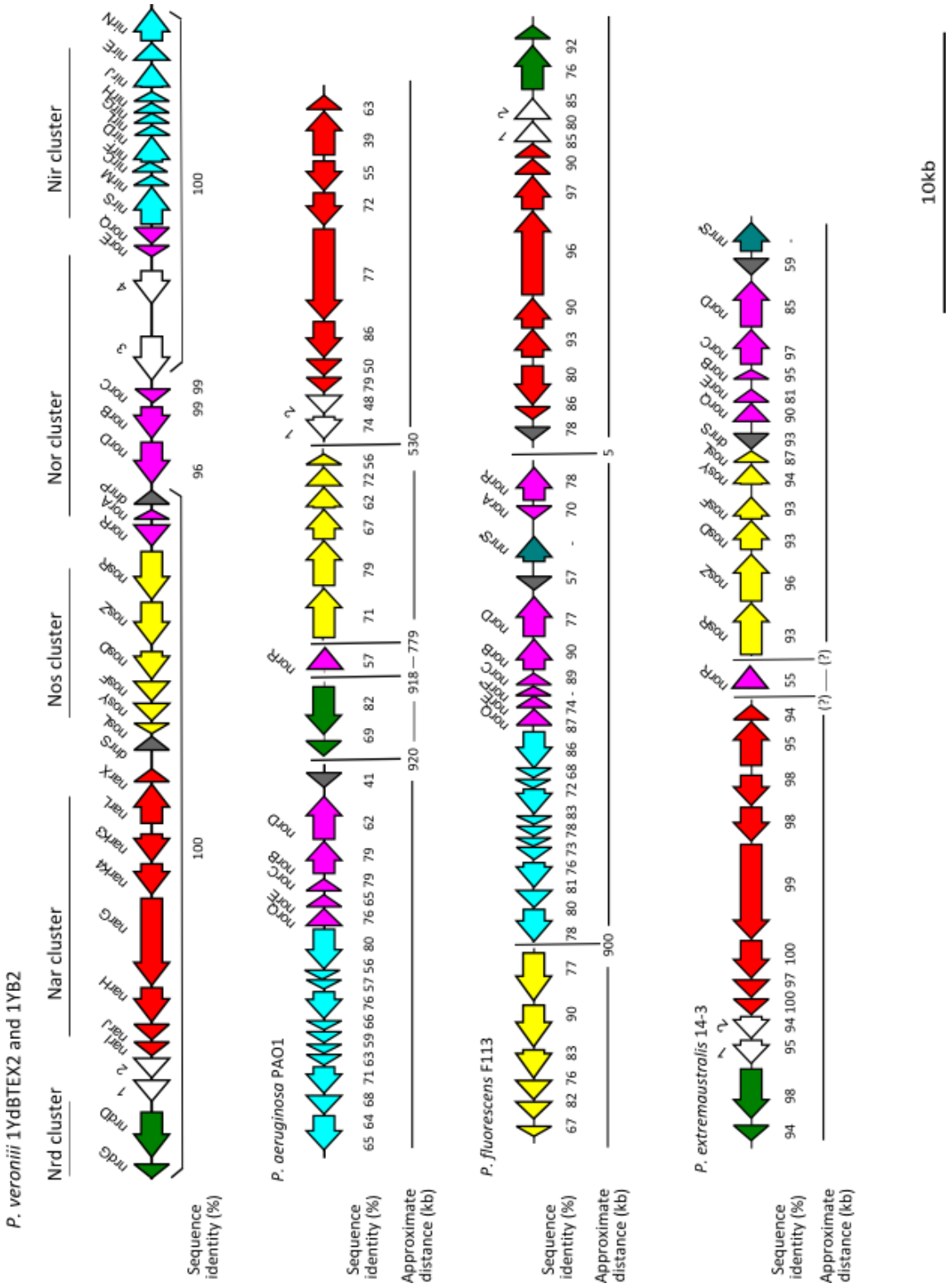


Figure 9: *P. veronii* 1YdBTEX2/1YB2, *P. aeruginosa* PAO1, *P. fluorescens* F113 and *P. extremaustralis* 14-3 denitrification gene clusters. Gene names are given above the 1YdBTEX2/1YB2 gene cluster. Only genes that are not in exactly the same locations as in 1YdBTEX2 and 1YB2 are named on the other strains. * indicates genes absent from *P. veronii* 1YdBTEX2 and 1YB2. Genes 1, 2, 3 and 4 were annotated as molybdenum cofactor biosynthesis protein MoaA, PpiC-type peptidyl-prolyl cis-trans isomerase, VgrG protein and Rhs-family protein respectively. The percent sequence identity between the encoded proteins of the *P. veronii* 1YdBTEX2 and 1YB2 gene clusters as well as those between the encoded proteins of *P.*

veronii 1YdBTEX2 on the one hand and *P. aeruginosa* PAO1, *P. fluorescens* F113 or *P. extremaustralis* 14-3 on the other hand are indicated below the gene clusters.

A gene cluster for the use of hydrogen as electron donor is localized in both P. veronii genomes

Besides the use of alternative electron acceptors, the use of alternative electron donors may by a further advantage for growth under environmental conditions. Various bacteria and archaea may use hydrogen as alternative electron donor and often use a [NiFe]-hydrogenase (Vignais et al., 2001). *Cupriavidus necator* H16 is one of the best characterized organisms for lithoautotrophic growth on molecular hydrogen and carbon dioxide (Friedrich and Schwartz, 1993) and hosts three distinct [NiFe]-hydrogenases which are encoded on megaplasmid pHG1 (Tran-Betcke et al., 1990; Kortlüke et al., 1992; Lenz and Friedrich, 1998; Schwartz et al., 2003). Under anaerobic conditions, the bacterium can switch to denitrification by using nitrate as alternative electron acceptor and hydrogen as electron donor (Römermann and Friedrich, 1985). Among *Pseudomonas* strains, the capability to use hydrogen as electron donor was not described and only recently a gene cluster encoding a [NiFe]-hydrogenase (Tribelli et al., 2012) was observed in *P. extremaustralis* 14-3. According to our knowledge this is the only *Pseudomonas* strain harboring those genes, although the functionality of the enzymes has not been published yet.

The nearly 20 kb gene clusters found in *P. veronii* 1YdBTEX2 and 1YB2 encode a [NiFe]-hydrogenase and the encoded proteins show a minimum of 97% sequence identity to each other. Comparison to the well-studied gene cluster from *C. necator* H16 (Schwartz et al, 2003) encoding a membrane bound [NiFe]-hydrogenases (MBH) show a high sequence identity of the encoded proteins of 42-89%.

In *C. necator* H16 this MBH gene cluster comprises genes encoding the structural enzymes (*hoxKGZ*), accessory proteins (*hoxMLOQRTV*) responsible for Ni-Fe center assembly (*hypABFCDEX*), a regulatory system (*hoxABCJ*) and a high affinity nickel transporter (*hoxN1*) (Schwarz et al., 2003). In *P. veronii* 1YdBTEX2 and 1YB2 the MBH gene cluster is of an organization similar to that in *C. necator* H16 (Schwartz et al, 2003), however, the regulatory genes precede those encoding structural proteins. In

addition *hypX* present in the H16 gene cluster is absent from both *P. veronii* 1YdBTEX2 and 1YB2.

The *hypX* gene encodes a maturation factor observed in aerobic hydrogen oxidizing microorganism, which in H16 was shown to confer oxygen tolerance on the soluble hydrogenase (Bleijlevens et al, 2004) but was also assumed to be required for the physiological function of the membrane bound hydrogenase under aerobic conditions (Schwartz et al., 2003), which may indicate the hydrogenase in *P. veronii* to be only functional under anaerobic conditions.

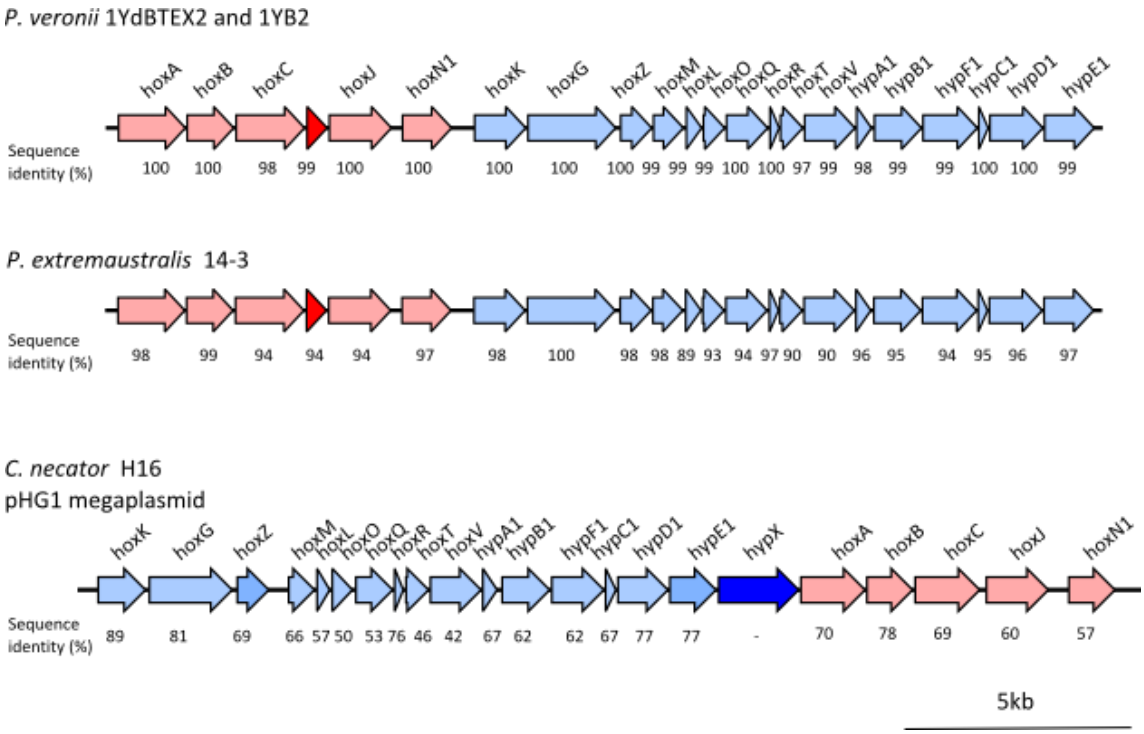


Figure 10: [NiFe]-hydrogenase encoding gene clusters of *P. veronii* 1YdBTEX2/1YB2, *P. extremaustralis* 14-3 and *C. necator* H16. Colored genes indicate the rearrangement in the different isolates. Gene colored in red was absent from H16 and *hypX,* colored in blue, was absent from strains 1YdBTEX2, 1YB2 and 14-3. The percent sequence identity between the encoded proteins of the *P. veronii* 1YdBTEX2 and 1YB2 gene clusters as well as those between the encoded proteins of *P. veronii* 1YdBTEX2 on the one hand and *C. necator* H16 or *P. extremaustralis* 14-3 on the other are given below the gene clusters.

4.5 Conclusions

Considering that *P. veronii* 1YdBTEX2 and 1YB2 were isolated from a contaminated benzene site, it was not surprising to find that they were able to degrade a variety of aromatic hydrocarbons. However, the use of at least three different catabolic gene clusters for benzene degradation including a novel benzene dihydrodiol dehydrogenase was not expected. This fact highlights the plasticity of the Pseudomonas genome in order to perform more efficient degradation, during dramatic modifications in the environment.

Both strains harbor additional traits, which can give an extraordinary competitive advantage under environmental conditions. In addition to the siderophore pyoverdin both *P. veronii* strains are capable of synthesizing a secondary siderophore pyochelin, which may add an advantage to thrive in iron limited environments. The presence of a complete denitrification pathway (NO_3^- to N_2) may allow them to grow under anaerobic conditions and the presence of a (NiFe) hydrogenase indicates that both *P. veronii* strains may use hydrogen as electron donor. Moreover, the use of nitrate as electron acceptor can be coupled to the use of H_2 as electron donor as described for *C. necator* H16 (Schwartz et al., 2003) however, in how far the described gene clusters encode functional enzymes remains to be elucidated.

4.6 Acknowledgements

This work was funded by BACSIN (project number: 211684) from the European Commission. Daiana de Lima Morales was supported by a grant from the CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (project number: 290020/2008-5). Special thanks to Dr. Robert Geffers from the sequencing facility at the Helmholtz Centre for Infection Research and to Iris Plumeier for technical assistance.

3.7 References

Aoki H, Kimura T, Habe H, Yamane H, Kodama T & Omori T (1996) Cloning, nucleotide sequence, and characterization of the genes encoding enzymes involved in the degradation of cumene to 2-hydroxy-6-oxo-7-methylocta-2,4-dienoic acid in *Pseudomonas fluorescens* IP01. *J Ferment Bioeng* **81***:* 187–196.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A & Zagnitko O (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Boetzer M, Henkel CV, Jansen HJ, Butler D & Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27***:* 578–9.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248–54.

Bray JR & Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* **27**: 325–349.

Cheuk W, Woo PCY, Yuen KY, Yu PH & Chan JKC (2000) Intestinal inflammatory pseudotumour with regional lymph node involvement: identification of a new bacterium as the aetiological agent. *J Pathol* **192**: 289–292.

Clarke KR & Warwick RM (2001) Change in marine communities: an approach to statistical analysis and interpretation, 2nd edition. PRIMER-E. *Plymouth*.

Cornelis P (2010) Iron uptake and metabolism in pseudomonads. *Appl Microbiol Biotechnol* **86***:* 1637–45.

Crosa JH & Walsh CT (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev* **66**: 223–249.

Darling A CE, Mau B, Blattner FR & Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394–403.

De Boer AP, van der Oost J, Reijnders WN, Westerhoff HV, Stouthamer AH & van Spanning RJ (1996) Mutational analysis of the nor gene cluster which encodes nitric-oxide reductase from *Paracoccus denitrificans*. *Eur J Biochem* **242**: 592–600.

De Lima-Morales D, Chaves-Moreno D, Jarek M, Vilchez-Vargas R, Jauregui R & Pieper DH (2013) Draft genome sequence of *Pseudomonas veronii* strain 1YdBTEX2. *Genome Announcs* **1**: 1–6.

Dorn E & Knackmuss HJ (1978) Chemical structure and biodegradability of halogenated aromatic compounds. Substituent effects on 1,2-dioxygenation of catechol. *Biochem J* **335**: 85–94.

Eaton RW (1996) p-Cumate catabolic pathway in *Pseudomonas putida* Fl: cloning and characterization of DNA carrying the *cmt* operon. *J Bacteriol* **178**: 1351–62.

Eaton RW, Selifonova OV & Gedney RM (1998) Isopropylbenzene catabolic pathway in *Pseudomonas putida* RE204: nucleotide sequence analysis of the *ipb* operon and neighboring DNA from pRE4. *Biodegradation* ***9***: 119–132.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Elomari M, Coroler L, Hoste B, Gillis M, Izard D & Leclerc H (1996) DNA relatedness among Pseudomonas strains isolated from natural mineral waters and proposal of *Pseudomonas veronii* sp. nov. *Int J Syst Bacteriol* **46**: 1138–1144.

Eltis LD & Bolin JT (1996) Evolutionary relationships among extradiol dioxygenases. *J Bacteriol* **178**: 5930–5937.

Friedrich B & Schwartz E (1993) Molecular biology of hydrogen utilization in aerobic chemolithotrophs. *Annu Rev Microbiol* **47**: 351–383.

Gaille C, Kast P & Haas D (2002) Salicylate biosynthesis *in Pseudomonas aeruginosa*. Purification and characterization of PchB, a novel bifunctional enzyme displaying isochorismate pyruvate-lyase and chorismate mutase activities. *J Biol Chem* **277**: 21768–21775.

Gaille C, Reimmann C & Haas D (2003) Isochorismate synthase (PchA), the first and rate-limiting enzyme in salicylate biosynthesis of *Pseudomonas aeruginosa*. *J Biol Chem* **278**: 16893–16898.

Gibson DT & Parales RE (2000) Aromatic hydrocarbon dioxygenases in environmental biotechnology. *Curr Opin Biotechnol* **11**: 236–243.

Heim S, Ferrer M, Heuer H, Regenhardt D, Nimtz M & Timmis KN (2003) Proteome reference map of *Pseudomonas putida* strain KT2440 for genome expression profiling: distinct responses of KT2440 and *Pseudomonas aeruginosa* strain PAO1 to iron deprivation and a new form of superoxide dismutase. *Environ Microbiol* **5**: 1257–1269.

Hernandez D, François P, Farinelli L, Osterås M & Schrenzel J (2008) De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res* **18**: 802–809.

Hong HB, Nam IH, Murugesan K, Kim YM & Chang YS (2004) Biodegradation of dibenzo-p-dioxin, dibenzofuran, and chlorodibenzo-p-dioxins by *Pseudomonas veronii* PH-03. *Biodegradation* **15**: 303–313.

Jousset A, Schuldes J, Keel C, Maurhofer M, Daniel R & Scheu S (2014) Full-genome sequence of the plant growth-promoting bacterium *Pseudomonas protegens* CHA0. *Genome Announc* **2**: 2014.

Junca H & Pieper DH (2010) Functional marker gene assays for hydrocarbon degrading microbial communities: Aerobic. In *Handbook of Hydrocarbon and Lipid Microbiology*. Timmis, KN (ed.). Berlin Heidelberg, Germany: Springer-Verlag, 4289-4312.

Junca H, Plumeier I, Hecht H-J & Pieper DH (2004) Difference in kinetic behaviour of catechol 2,3-dioxygenase variants from a polluted environment. *Microbiology* **150**: 4181–4187.

Kortlüke C, Horstmann K, Schwartz E, Rohde M, Binsack R & Friedrich B (1992) A gene complex coding for the membrane-bound hydrogenase of *Alcaligenes eutrophus* H16. *J Bacteriol* **174:** 6277–6289.

Kupferschmied P, Maurhofer M & Keel C (2013) Promise for plant pest control: root-associated pseudomonads with insecticidal activities. *Front Plant Sci* **4***:* 287.

Laemmli UK (1970) Cleavage of Structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Lenz O & Friedrich B (1998) A novel multicomponent regulatory system mediates H2 sensing in *Alcaligenes eutrophus*. *Proc Natl Acad Sci U S A* **95**: 12474–12479.

Loper JE, Hassan KA, Mavrodi DV, Davis EW 2nd, Lim CK, Shaffer BT, Elbourne LD, Stockwell VO, Hartney SL, Breakwell K, Henkels MD, Tetu SG, Rangel LI,Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS 3rd, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE & Paulsen IT (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* **8**: e1002784.

Martin VJ & Mohn WW (1999) A novel aromatic-ring-hydroxylating dioxygenase from the diterpenoid-degrading bacterium *Pseudomonas abietaniphila* BKME-9. *J Bacteriol* **181**: 2675–2682.

Martin VJ & Mohn WW (2000) Genetic investigation of the catabolic pathway for degradation of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. *J Bacteriol* **182**: 3784–3793.

Moreno-Hagelsieb G & Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**: 319–324.

Mulet M, Lalucat J & García-Valdés E (2010) DNA sequence-based analysis of the Pseudomonas species. *Environ Microbiol* **12**: 1513–1530.

Nam I-H, Chang Y-S, Hong HB & Lee YE (2003) A novel catabolic activity of *Pseudomonas veronii* in biotransformation of pentachlorophenol. *Appl Microbiol Biotechnol* **62**: 284–290.

Onaca C, Kieninger M, Engesser K-H & Altenbuchner J (2007) Degradation of alkyl methyl ketones by *Pseudomonas veronii* MEK700. *J Bacteriol* **189**: 3759–3767.

Ortet P, Barakat M, Lalaouna D, Fochesato S, Barbe V, Vacherie B, Santaella C, Heulin T & Achouak W (2011) Complete genome sequence of a beneficial plant root-associated bacterium, *Pseudomonas brassicacearum*. *J Bacteriol* **193**: 3146.

Patel HM & Walsh CT (2001) In vitro reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyochelin: Characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. *Biochemistry* **40**: 9023–9031.

Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GS, Mavrodi DV, DeBoy RT, Seshadri R, Ren Q, Madupu R, Dodson RJ, Durkin AS, Brinkac LM, Daugherty SC, Sullivan SA, Rosovitz MJ, Gwinn ML, Zhou L, Schneider DJ, Cartinhour SW, Nelson WC, Weidman J, Watkins K, Tran K, Khouri H, Pierson EA, Pierson LS 3rd, Thomashow LS & Loper JE (2005) Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nat Biotechnol* **23**: 873–878.

Philippot L (2002) Denitrifying genes in bacterial and archaeal genomes. *Biochim Biophys Acta* **1577**: 355–376.

Philippot L, Mirleau P, Mazurier S, Siblot S, Hartmann A, Lemanceau P & Germon JC (2001) Characterization and transcriptional analysis of *Pseudomonas fluorescens* denitrifying clusters containing the *nar*, *nir*, *nor* and *nos* genes. *Biochim Biophys Acta* **1517**: 436–440.

Pohlmann A, Cramm R, Schmelz K & Friedrich B (2000) A novel NO-responding regulator controls the reduction of nitric oxide in *Ralstonia eutropha*. *Mol Microbiol* **38**: 626–638.

Pukatzki S, Ma AT, Revel AT, Sturtevant D & Mekalanos JJ (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc Natl Acad Sci U S A* **104**: 15508–15513.

Quadri LEN, Keating TA, Patel HM & Walsh CT (1999) Assembly of the *Pseudomonas aeruginosa* nonribosomal peptide siderophore pyochelin: In Vitro reconstitution of aryl-4,2-bisthiazoline synthetase activity from PchD, PchE, and PchF. *Biochemistry* **38**: 14941–14954.

Ramette A, Frapolli M, Saux MF-L, Gruffaz C, Meyer J-M, Défago G, Sutra L & Moënne-Loccoz Y (2011) *Pseudomonas protegens* sp. nov., widespread plant-protecting bacteria producing the biocontrol compounds 2,4-diacetylphloroglucinol and pyoluteorin. *Syst Appl Microbiol* **34**: 180–188.

Redondo-Nieto M, Barret M, Morrisey JP, Germaine K, Martínez-Granero F, Barahona E, Navazo A, Sánchez-Contreras M, Moynihan JA, Giddens SR,Coppoolse ER, Muriel C, Stiekema WJ, Rainey PB, Dowling D, O'Gara F, Martín M & Rivilla R (2012) Genome sequence of the biocontrol strain *Pseudomonas fluorescens* F113. *J Bacteriol* **194**: 1273–1274.

Reimmann C (2012) Inner-membrane transporters for the siderophores pyochelin in Pseudomonas aeruginosa and enantio-pyochelin in *Pseudomonas fluorescens* display different enantioselectivities. *Microbiology* **158**: 1317-1324.

Reimmann C, Patel HM, Serino L, Barone M, Walsh CT & Haas D (2001) Essential PchG-dependent reduction in pyochelin biosynthesis of *Pseudomonas aeruginosa*. *J Bacteriol* **183**: 813-820.

Romermann D & Friedrich B (1985) Denitrification by *Alcaligenes eutrophus* is plasmid dependent. *J Bacteriol* **162**: 852–854.

Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Schwartz E, Henne A, Cramm R, Eitinger T, Friedrich B & Gottschalk G (2003) Complete nucleotide sequence of pHG1: A *Ralstonia eutropha* H16 megaplasmid encoding key enzymes of H2-based lithoautotrophy and anaerobiosis. *J Mol Biol* **332**: 369–383.

Serino L, Reimmann C, Visca P, Beyeler M, Chiesa, V Della & Haas D (1997) Biosynthesis of pyochelin and dihydroaeruginoic acid requires the iron-regulated *pch*DCBA operon in *Pseudomonas aeruginosa*. *J Bacteriol* **179**: 248-257.

Siedow A, Cramm R, Siddiqui RA & Friedrich B (1999) A megaplasmid-borne anaerobic ribonucleotide reductase in *Alcaligenes eutrophus* H16. *J Bacteriol* **181**: 4919–4928.

Sierra R, Rodríguez-R LM, Chaves D, Pinzón A, Grajales A, Rojas A, Mutis G, Cárdenas M, Burbano D, Jiménez P, Bernal A & Restrepo S (2010) Discovery of Phytophthora infestans genes expressed in planta through mining of cDNA libraries. *PloS One* **5**: e9847.

Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, Knight CG, Malone JG,Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS, Bentley SD, Hothersall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB & Thomson NR (2009) Genomic and genetic analyses of diversity and plant interactions of P*seudomonas fluorescens*. *Genome Biol* **10**: R51.

Silby MW, Winstanley C, Godfrey SAC, Levy SB & Jackson RW (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev* **35**: 652–680.

Smith DJ, Martin VJJ & Mohn WW (2004) A Cytochrome P450 Involved in the metabolism of abietane diterpenoids by *Pseudomonas abietaniphila* BKME-9. *J Bacteriol* **186**: 3631–3639.

Smith DJ, Park J, Tiedje JM & Mohn WW (2007) A large gene cluster in Burkholderia xenovorans encoding abietane diterpenoid catabolism. *J Bacteriol* **189**: 6195–6204.

Smith DJ, Patrauchan MA, Florizone C, Eltis LD & Mohn WW (2008) Distinct roles for two CYP226 family cytochromes P450 in abietane diterpernoid catabolism by Burkholderia xenovorans LB400. *J Bacteriol* **190**: 1575–1583.

Sommer DD, Delcher AL, Salzberg SL & Pop M (2007) Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* **8**: 64.

Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E,Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH,Hancock RE, Lory S & Olson MV (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**: 959–964.

Takeo M, Prabu SK, Kitamura C, Hirai M, Takahashi H, Kato D-I & Negoro S (2006) Characterization of alkylphenol degradation gene cluster in *Pseudomonas putida* MT4 and evidence of oxidation of alkylphenols and alkylcatechols with medium-length alkyl chain. *J Biosci Bioeng* **102**: 352–361.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Tran-Betcke A, Warnecke U, Böcker C, Zaborosch C & Friedrich B (1990) Cloning and nucleotide sequences of the genes for the subunits of NAD-reducing hydrogenase of *Alcaligenes eutrophus* H16. *J Bacteriol* **172**: 2920–2929.

Tribelli PM, Raiger Iustman LJ, Catone MV, Di Martino C, Revale S, Méndez BS & López NI (2012) Genome sequence of the polyhydroxybutyrate producer *Pseudomonas extremaustralis*, a highly stress-resistant Antarctic bacterium. *J Bacteriol* **194**: 2381–2382.

Vignais PM, Billoud B & Meyer J (2001) Classification and phylogeny of hydrogenases 1. *FEMS Microbiol Rev* **45**: 455-501.

Vollack KU, Härtig E, Körner H & Zumft WG (1999) Multiple transcription factors of the FNR family in denitrifying *Pseudomonas stutzeri*: characterization of four *fnr*-like genes, regulatory responses and cognate metabolic processes. *Mol Microbiol* **31**: 1681–1694.

Vullo DL, Ceretti HM, Daniel MA, Ramírez SAM & Zalts A (2008) Cadmium, zinc and copper biosorption mediated by *Pseudomonas veronii* 2E. *Bioresour Technol* **99**: 5574–5581.

Witzig R, Aly HAH, Strömpl C, Wray V, Junca H & Pieper DH (2007) Molecular detection and diversity of novel diterpenoid dioxygenase DitA1 genes from proteobacterial strains and soil samples. *Environ Microbiol* **9**: 1202–1218.

Witzig R, Junca H, Hecht HJ & Pieper DH (2006) Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: Links between benzene

biodegradation and genes similar to those encoding isopropylbenzene dioxygenases. *Appl Environ Microbiol* **72**: 3504–3514.

Yamamoto S & Harayama S (1995) PCR amplification and direct sequencing of *gyr*B genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* **61**: 3768.

Yen KM & Gunsalus IC (1982) Plasmid gene organization: naphthalene/salicylate oxidation. *Proc Natl Acad Sci U S A* **79**: 874–878.

Youard ZA, Mislin GLA, Majcherczyk PA, Schalk IJ & Reimmann C (2007) *Pseudomonas fluorescens* CHA0 produces enantio-pyochelin, the optical antipode of the *Pseudomonas aeruginosa* siderophore pyochelin. *J Biol Chem* **282**: 35546–35553.

Zachow C, Alavi M, Tilcher R & Berg G (2013) Complete genome sequence of the sugar beet endophyte *Pseudomonas poae* RE*1-1-14, a disease-suppressive bacterium. *Genome Announc* **1**: 2–3.

Zerbino DR & Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18***:* 821–829.

Chapter IV – Microbial community and catabolic gene structure in soil under benzene and BTEX stress

Daiana Lima-Morales[1], Ruy Jáuregui[1], Amelia Camarinha-Silva[1*], Ramiro Vilchez-Vargas[1], Dietmar H. Pieper[1]**

[1]Microbial Interactions and Processes Research Group, Department of Medical Microbiology, Helmholtz Centre for Infection Research, Braunschweig, Germany; [2] Genome Analytics, Helmholtz Centre for Infection Research, Braunschweig, Germany; [*] Present address Institute of Animal Nutrition, University of Hohenheim, Stuttgart, Germany, [**] Present address Laboratory of Microbial Ecology and Technology (LabMET), University of Ghent, Ghent, Belgium

Keywords: Illumina deep-sequencing, BTEX, microarray, bacterial community

Running title Microbial community under BTEX stress

Microbial community and catabolic gene structure in soil under benzene and BTEX stress

4.1 Abstract

In order to identify commonalities or differences in the adaptation of microbial communities, three types of soil were subjected to a constant supply of benzene or benzene/toluene/ethylbenzene/xylenes for a period of 3 month. Whereas soil previously subjected to intensive *in-situ* bioremediation showed only negligible changes in community structure, other contaminated soil samples showed a clear succession of phylotypes. A rapid response to benzene stress was observed whereas the response to BTEX pollution was significantly slower. After extended incubation, actinobacterial phyloypes were increasing in relative abundance, indicating their superior fitness to pollution stress. Commonalities, but also differences in the phylotypes were observed. Catabolic gene surveys confirmed the enrichment of actinobacteria by identifying the increase of actinobacterial genes involved in the degradation of hydrophobic pollutants, as well as of catabolic genes, which can serve as phylogenetic markers. Proteobacterial phylotypes were increasing in relative abundance in one of the soils after short term stress with benzene, where catabolic gene surveys indicated metabolic routes enriched. Interestingly, soil having been subjected to intensive *in-situ* bioremediation, despite staying constant in community structure, showed an adaptation in the metabolic net, indicating that a highly adapted community has been enriched, which had to adapt its gene pool to novel challenges.

4.2 Introduction

BTEX (benzene, toluene, ethylbenzene and the isomers of xylene) are of major concern for human health, and are classified as priority pollutants (Eriksson et al., 1998). It is primordial to avoid that these chemicals enter the environment. Nevertheless losses of contaminants during industrial and commercial operations, municipal and industrial waste treatment, oil extraction and derivatives production, retail distribution of petro products, inadequate storage and sale are the main sources of BTEX environment contamination (Coseuil & Marins, 1997; Shim et al., 2001) (http://www.gsi-net.com/files/papers/meamodel.pdf). In addition BTEX can enter the environment by huge disasters like the recent *Deepwater Horizon oil spill* where approximately 4.9 million barrels of crude oil entered the deep ocean (http://www.restorethegulf.gov/sites/default/files/documents/pdf/OilBudgetCalc_Full_HQ-Print_111110.pdf).

Several microorganisms have evolved specialized pathways to use aromatic compounds like BTEX as their sole carbon and energy source (Pérez-Pantoja et al., 2009). The analysis of aromatic degradation by isolates gives a valuable structured understanding of metabolic pathways, where the key steps are the ring activation and the ring cleavage. Some important monoaromatic degradation pathways described are the TOD pathway of *P. putida* F1, where the aromatic ring is activated by a dioxygenase (encoded by the *todC1C2BA* genes) and cleaved by a catechol 2,3-dioxygenase (encoded by the *todE* gene) (Gibson et al., 1968a, 1968b), the TOM pathway of *B. cepacia* G4, where the aromatic ring is activated by two hydroxylations catalyzed by toluene 2-monooxygenase (encoded by the *tomA0A1A2A3A4A5* genes) (Newman and Wackett 1995) and the catechol is cleaved by a catechol 2,3-dioxygenase (Newman and Wackett, 1995) or the TOL pathway encoded on plasmid pWW0 of *P. putida* mt2, where the degradation is initiated by the oxidation of the methyl substituent by a xylene monooxygenase (encoded by the *xyl*MA genes) and the catechol structure is cleaved by a catechol 2,3-dioxygenase (encoded by the *xyl*E gene) (Williams and Murray, 1974; Greated et al., 2002). However, even though a huge set of information has been generated using microorganisms enriched in the laboratory, it is

known they often don't play an important role in *in situ* biodegradation of pollutants (Jeon et al., 2003; Witzig et al., 2006).

Eliminating the need of cultivation, molecular techniques allow studying the microbial diversity and activity in situ. Some reports have focused on the molecular characterization of BTEX degradation genes (Junca & Pieper, 2004; Hendrickx et al., 2006; Nebe et al., 2009) through clone libraries or molecular fingerprinting methods and it was shown that catabolic genes selected often differ from those expected to be enriched taking into account culture-based information (Witzig et al., 2006). However the described molecular techniques are limited by the primer selection for amplicon preparation and, due to the high diversity of catabolic genes catalyzing a given reaction, (Perez-Pantoja et al., 2009) they can only cover a single gene subfamily and not the whole diversity.

The development of functional gene arrays solves the limitations of former fingerprinting techniques. The microarrays are a powerful genomic technology that generate massive information from single organisms but also from environmental samples. Among the environmental oligoarrays, the different versions of the so called GeoChip (Berthrong et al., 2009; He et al., 2010) received special attention (He et al., 2007; Wang et al., 2009; He et al., 2010; Van Nostrand et al., 2011). More recently a catabolome array was described (Vilchez-Vargas et al., 2013), which is based on upgraded databases of key aromatic catabolic genes (Perez-Pantoja et al., 2009) and avoids the massive amounts of misannotations in proteins where probes were designed on the Geochip. The array comprises 1500 probes, where the design facilitates to assign positive signals to the respective protein subfamilies, therefore directly inferring function and substrate specificity.

Recent studies have characterized microbial communities from contaminated environments, using "next generation sequencing" (NGS) high throughput technologies. The pyrosequencing via the Roche 454 FLX instrument with Titanium reagents has been applied in many recent studies to analyze the bacterial diversity in the *Deepwater Horizon oil spill* (Mason et al., 2012), in oil mousses, sediments and seawater (Liu & Liu 2013), in bioreactors supplemented with polycyclic aromatic hydrocarbons (Singleton et al. 2011) or in diesel contaminated soil (Sutton et al., 2012)

among others. A handful of publications have now assessed the applicability of using the Illumina technology for amplicon deep-sequencing, and reported its advantages and shortcomings over other technologies, particularly in relation to error-rates, the need for quality filtering and taxonomic classification (Caporaso et al., 2012; Degnan & Ochman, 2012; Soergel et al., 2012; Werner et al., 2012; Bokulich et al., 2013). To analyze which microorganisms are important in BTEX contaminated soils, we upgraded a high throughput method using the Illumina platform (Camarinha-Silva et al., 2014) using the V5/V6 hypervariable region of the 16S rRNA gene.

Studying complex community structures and their involvement in bioremediation is challenging and requires a multifaceted approach. The era of NGS in particular deep-sequencing via the Illumina platform and the microarray technology have provided us with the necessary tools to study the communities as a whole. The aim of this work is to analyze microbial community structure and catabolic gene diversity in contaminated soils of different origins under constant and specific contamination with benzene and BTEX in microcosm experiments over time and to identify shifts in community structure and prevalence of key genes.

4.3 Materials and Methods

4.3.1 Samples

Three different contaminated soil samples were used in this study. The czech soil sample (CZE) was provided by *AECOM CZ s.r.o.*, Prague, Czech Republic and derived from a former army airbase Hradcany (CZE). The site was contaminated with petroleum hydrocarbons (mainly by jet fuel). This soil passed through a 3-year clean-up and exhibits a moderate organic contamination (Kabelitz et al., 2009). A second soil sample was provided by Belair Biotech S.A., Geneva, Switzerland. It was collected in a soil treatment station, in Geneva city and it was previously contaminated with industrial waste (SUI). The third soil was provided by Braskem S.A., it was collected in the south of Brazil, in a petrol station, where the ground soil was contaminated with fuel, mainly gasoline (BRA). For more information on soil characteristics see Table 1.

Table 1: Chemical and physical properties of soil samples.

		CZE	BRA	SUI
pH		6.4	4,89	7,62
Texture				
% Sand (53µm)	%	83.7	57,8	62,5
% Clay	%	1.67	18,4	7,9
% Silt	%	ND	17,6	10,3
Total C	%	ND	0,395	2,4483
Total N	%	ND	0,0436	0,0873
Organic C	%	ND	0,4447	0,4577
$CaCO_3$	%	ND	3,71	25,34
CEC	cmol/kg	ND	10,27	9,52
Mg	cmol/kg	ND	1,82	0,45
K	cmol/kg	ND	0,11	0,07
Ca	cmol/kg	ND	4,41	18,13

ND, not determined

4.3.2 Experiment design

Soil samples (4 g) were deposited in Erlenmeyer flasks (150 ml total volume), where benzene or a BTEX mixture were supplied via the vapor phase by evaporation from a glass tube placed through the screw caps closing the Erlenmeyer flasks. Benzene, toluene, ethylbenzene and xylene isomers were obtained from Sigma-Aldrich Chemie (Steinheim, Germany). The microcosms were incubated at 15°C. The amount of contaminants inside the flasks, was kept constant during the whole experiment.

To do this, either benzene (26,6 µl) or BTEX (26,6 µl of each benzene, toluene and ethylbenzene and 8,86 µl of each xylene isomer) were added to the glass tubes in a volume of 500 µl heptamethylnonane. The amount of pollutant in the gas phase corresponded to 0,7 mM of benzene, 0,2 mM of toluene, 0,2 mM of total xylenes and 0,1 mM of ethylbenzene as compared to authentic standards. Due to evaporation, glass tubes were exchanged all 4 days.

BTEX components in gas phase were determined by Capillary GC, on a 6890N GC from Agilent Technologies gas chromatograph equipped with a capillary column Optima5 (5% diphenylpolysiloxane 95% dimethylpolysiloxane; length, 50 m; inner diameter of 0,32 mm and 0,25µm film thickness). Carrier gas was hydrogen with a flow of 70 ml/min. The oven temperature was maintained at 60°C for 1 min, then increased 3°C/min until 80°C and finally 25°C/min until 150°C. A 10 µL gas tight syringe was used to inject 2.5 µL volume directly through the GC septum. The injector temperature was set at 220°C and the detector temperature at 300°C (FID).

The microcosms were incubated at 15°C for a period of 90 days and samples taken after 2, 4, 10, 20, 30, 60 and 90 days of incubation. At each time point a whole microcosms was harvested and DNA was extracted from 2 g of soil. Original soil (2 g) was used as control (time 0).

4.3.3 Catabolic gene analysis

Genomic DNA analysis using the catabolic array

The catabolic array consisted of a total of 1500 probes spotted in triplicate across the glass surface as previously described (Vilchez-Vargas et al. 2013).

DNA extraction

DNA was extracted according to Martin-Laurent et al. (2001). Briefly, 2 g of soil were mixed with 1 ml Tris/HCl (100 mM pH 8.0), supplemented with 100 mM EDTA, 100 mMNaCl, 1% (wt/vol) polyvinylpyrrolidone and 2% (wt/vol) sodium dodecyl sulfate and transferred into a 2 ml Lysing Matrix E tube (Qbiogene, Alexis® Biochemicals, Carlsbad, CA). Cells were lysed in a Fast Prep®-24 instrument (40 sec, 6.0 m/sec). Samples were centrifuged at 14.000 x g for 1 min at room temperature and the supernatant washed with 1 volume phenol/chloroform (1:1), centrifuged and the aqueous phase washed with 1 volume chloroform. After centrifugation, nucleic acids (aqueous phase) were precipitated with 1 volume of ice-cold isopropanol and 1:10 volume of 3M sodium acetate and purified using a Sepharose 4B (Sigma-Aldrich) spin column (MP

Biomedicals). Typically, three-four column purification steps were necessary to reach a transparent solution, from which DNA was precipitated with 3 volumes of ethanol and 1:10 volume of 3 M sodium acetate. After centrifugation and washing with 80% ethanol the pellet was resuspended in 20 µl of milliQ water. The quality and quantity of the DNA samples were analyzed on 1% agarose gels and spectrophotometrically by determination of the A260/A280 ratios.

Array sample preparation

DNA amplification was performed with 100 ng of DNA using the REPLI-g Ultrafast kit (Qiagen) following the instructions of the manufacturer, however, with a doubled reaction volume (40 µl) and incubation at 30°C for 2.5 h. Amplified DNA was purified by phenol/chloroform extraction and DNA precipitated with 3 volumes of ethanol and 1:10 volume of 3 M sodium acetate. After centrifugation the pellet was rinsed twice with 80% ethanol, suspended in 100 µl of milliQ water and heat-fragmented at 95°C for up to 1.5 h. The digestion was followed by analysis of aliquots on 1.5% agarose gels and terminated as soon as the majority of DNA fragments had a size of 200 - 1000 bp. After precipitation, DNA was suspended in 45 µl of milliQ water and labeled using terminal transferase (Roche) and Cy5-dUTP (GE Healthcare) following the manufacturers' instructions although with incubation at 37 °C for 4 h. Labeled DNA-fragments were precipitated and suspended in 20 µl of milliQ water, incubated with herring sperm DNA (1 µl, corresponding to 10 µg, Invitrogen) for 5 min at 95°C and 80 µl of hybridization buffer (HB) added.

For hybridization, slides (CodeLink Activated Slides, SurModics, Eden Praine, MN, USA) were inserted into the hybridization chamber (Slide-Booster SB401/800, Advalytics Beckman Coulter Biomedical GmbH, Munich, Germany) and covered by coverslips (Implen GmbH, Munich, Germany). Samples were added by diffusion. DNA was hybridized at 55°C during 18 h, using hybridization buffer comprising 15% (v/v) dimethylsulfoxide, 25% (v/v) formamide, 1.25x SSC (190 mM sodium chloride plus 20 mM tri-sodium citrate equivalent to a sodium concentration of 250 mM), 0.15% SDS, 0.15% Tween20, 880 mM Betaine, 5x TE buffer (50 mM Tris-HCl, 5 mM EDTA) and 0.1 mg ml-1 BSA in aqueous solution. This corresponds to a final concentration in the

hybridization mixture of 1x SSC. The slides were subsequently washed in 1x SSC 0.3% SDS (5 min, 42°C), twice in 1xSSC (1 min each, 20°C), in 0.5xSSC (1 min,20°C), in 0.1x SSC 0.3% SDS (1 min, 42°C) and twice in 0.1x SSC (1 min, 20°C). Slides were dried by centrifugation (ArrayIt Microarray High-Speed Centrifuge, Arrayit Corporation, Sunnyvale, CA, USA) and scanned using a High Resolution Microarray Scanner (Agilent Technologies, Life Sciences and Chemical Analysis Group, Santa Clara, CA).

Array sample analysis

All samples were analysed for their catabolic gene structure by the microarray. After scanning, spots corresponding to the internal controls 50 mer A–D were used to generate a standard curve (Vilchez-Vargas et al., 2013) and those samples where a linear curve could not be observed were discarded and 35 out of 63 microarrays were further analysed. Signal Intensities were normalized against the background using the formula NI=(probe intensity-background intensity)/background intensity and the average of the three experimental replicas determined.

4.3.4 Community analysis

16S rDNA amplicon preparation

The V5 and V6 regions of the 16S rRNA gene were amplified using primers 807F and 1050R (Table 2) with 1 µL soil DNA as template in a total volume of 20 µL 5x PrimeSTAR TM buffer, containing each deoxynucleoside triphosphate at a concentration of 2.5 mM, each primer at a concentration of 0.2 µM, 1 µL of template DNA and 0.2 µL of PrimeSTAR TM HS DNA polymerase (2.5U). An initial denaturation step of 95°C for 3 min was followed by 20 cycles of denaturation at 98°C for 10 sec, annealing at 51°C for 10 sec and extension at 72°C for 45 sec. One µL of the first reaction mixture served as template in a second PCR performed under the same conditions as above described, but for 15 cycles, where the 807F forward primer contains a 6 nucleotides (nt) error correcting barcode (Meyer &Kircher, 2010). Both primers comprised sequences complementary to the illumina specific adaptors to the 5´ends. In a third amplification reaction (10 cycles), 1 µL of the second reaction

mixture was used as template using PCR primers designed to integrate the sequence of the specific Illumina multiplexing sequencing primers and index primers. PCR amplicons were verified by agarose gel electrophoresis, purified using Macherey-Nagel 96-well plate purification kits (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions and quantified with the Quant-iTPicoGreendsDNA reagent and kit (Invitrogen). Libraries were prepared by pooling equimolar ratios of amplicons (200 ng of each sample) derived from each time sample from each soil, all having been tagged with a unique barcode. To remove single nucleotides and concentrate the sample, each library (626–1169 µl of volume) was precipitated on ice for 30 min after addition of 20 µl of NaCl (3M) and 3 volumes of ice-cold 100% ethanol. The precipitated DNA was centrifuged at 13,000 *x* g for 30 min at 4°C. The supernatant was removed, the pellet air dried, resuspended in 30 µL of double-distilled water and separated on a 2% agarose gel. PCR products of the correct size were extracted and recovered using the QIAquick gel extraction kit (Qiagen). Negative controls using water as template were performed and were free of any amplification products after all rounds of PCR.

Table 2: Primers used for the amplications of V5 and V6 regions of the 16S rRNA gene.

Primer name	Primer Sequence (5'-3')	Reference
807 Fw	GGATTAGATACCCBRGTAGTC	Bohorquez et al., 2012
1050 Rev	AGYTGDCGACRRCCRTGCA	
Illu5-6 For 1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGATTGTGGATTAGATACCCBRGTAGTC	This study
Illu5-6 For 2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTTAATGGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 3	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCAGGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 4	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGACCGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 5	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAATGGAGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 6	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACGTTGGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 7	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCTCGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 8	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAACTTGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 9	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGTAAGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 10	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTAACGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 11	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGAGAGGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 12	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCATAGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 13	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTCCAGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 14	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTATGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 15	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACGCAGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 16	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACTGCGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 17	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTTGGGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 18	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAACTGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 19	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGACGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 20	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATATTGGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 21	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAAGAGGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 22	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAATTCGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 23	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGGCCGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 24	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAACGGTGGATTAGATACCCBRGTAGTC	
Illu5-6 For 25	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTCCGGTGGATTAGATACCCBRGTAGTC	
Illu Rev Adapter V5-V6	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGYTGDCGACRRCCRTGCA	
Multiplexing PCR primer 1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	
Index 1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTC	Illumina Inc.
Index 2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTC	
Index 4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTC	
Index 6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTC	
Index 8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTC	

4.3.5 Bioinformatic analysis

For this study, only the forward end sequence reads were processed. In total 2.003.786 sequence reads were obtained. A quality filter program that runs a sliding window over the read and calculates the local average score based on the illumina quality line of the fasta file, trimmed 3'-ends of the reads that fall below a quality score of 10 (http://bioinformatics.ucdavis.edu/index.php/Trim.pl). Only reads of a minimum of 115 nt in length (29 nt of primer and barcode sequence and 86 nt of 16S rRNA gene sequence) were further analyzed. All truncated reads that had an N character in their sequence, any mismatches within primers and barcodes or more than 8 homopolymer stretches were discarded. All sequences from each sample were split into different files according to their unique barcode.

A total of 63 samples comprising >3000 sequence reads each were further processed, totaling to 1,671,472 sequence reads. This data-set was collapsed into unique representative reads. These reads were pre-clustered into 24,000 unique representative reads using a pseudo-single-linkage algorithm which removes sequences most probably resulting from sequencing errors (Schloss et al., 2009). A representative read was further considered if: a) it was present in at least one sample in a relative abundance >0,1% of the total sequences of that sample or b) it was present in at least 3 samples or c) it was present in a copy number of at least 10 in at least one sample. Phylotype representatives were then generated by clustering at 98% similarity (1 mismatch) using USEARCH (Edgar, 2010). This reduced the number of representative reads to 501 phylotypes, a computational manageable level without curtailing the fine scale soil community composition. All phylotypes were assigned a taxonomic affiliation based on naive Bayesian classification (RDP classifier) (Wang et al., 2007). All 501 phylotypes were then manually analyzed against the RDP database using the Seqmatch function as well as against the NCBI database to define the discriminatory power of each sequence read.

A species name was assigned to a phylotype when only 16S rRNA gene fragments of previously described isolates of that species showed ≤ 2 mismatches with the respective representative sequence read. Similarly, a genus name was assigned to a phylotype when only 16S rRNA gene fragments of previously described isolates belonging to that genus and of 16S rRNA gene fragments originating from uncultured representatives of that genus showed ≤ 2 mismatches.

4.4 Results and Discussion

Three different soils from the Czech Republic (CZE), Brazil (BRA) and Switzerland (SUI) were subjected to pollutant stress with either benzene or BTEX and compared to a control. Bacterial community shifts were analyzed by Next Generation Sequencing of the 16S rRNA gene and catabolic gene shifts were analyzed by a catabolic gene array (Vilchez-Vargaz et al., 2013).

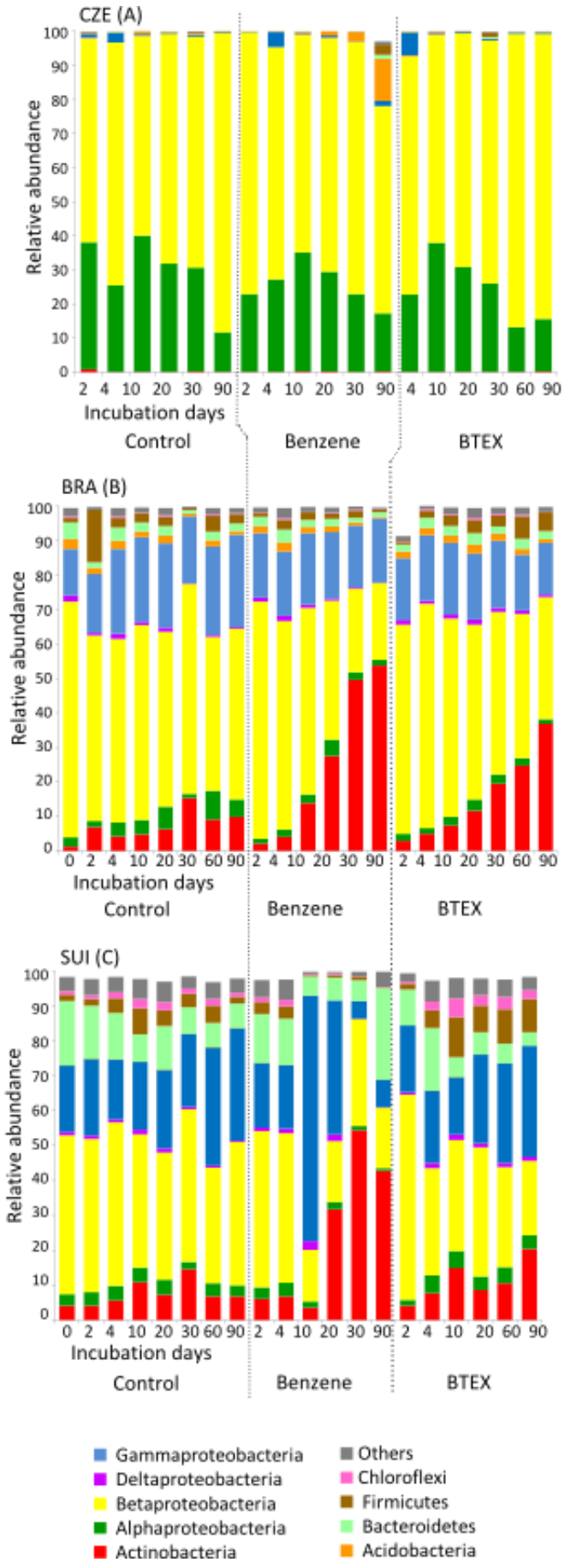
4.4.1 Soil community structure

Analysis of 86 bases of the V5 hypervariable region of the 16S rRNA gene indicated the presence of a total of 501 phylotypes, out of which 470 could be identified to the phylum level, 432 to the class level, 361 to the order level and 312 to the family level. 187 phylotypes were observed in the BRA soil, 304 in the soil from SUI, but only 58 in the CZE soil (Table S1).

Regarding the class/phylum level composition, the BRA and the SUI soils are dominated by β- and γ-Proteobacteria. Members of these classes were also observed in the CZE soil, however, specifically γ-Proteobacteria were present in lower abundance. The CZE soil differ from the other soils by its low diversity, where members of the α- and β-Proteobacteria, accounted for more than 97% of sequence reads (Figure 1).

Changes in community and catabolic gene structure during benzene and BTEX stress in CZE soil

Analysis of the community structure over time showed only a slight effect of the incubation conditions on the community structure. After 10 days of incubation an increase of one phylotype (Phy108) indicating the presence of Acidobacteria was observed in the benzene treated microcosms. After 90 days this unique phylotype reached >10% of the total abundance (Figure 1A).



According to molecular studies, Acidobacteria are among the most abundant bacterial phyla in soil (Jansse, 2006) and may represent up to 20% of total bacteria in soil communities (Nacke et al., 2011; Naether et al., 2012). However, due to the difficulty to culture them, little is known about members of this phylum. Recent genome sequencing projects targeting bacteria belonging to this group promoted insights into their life-style. Ward et al. (2009) suggested that Acidobacteria are best suited to low-nutrient conditions and are resistant to desiccation. Strain *Holophaga foetida* TMBS4T was even capable to degrade several aromatic compounds such as gallate, phloroglucinol, or pyrogallol under anaerobic conditions (Liesack et al., 1994; Anderson et al., 2012) and Acidobacteria have also been identified as abundant in a gasoline spill (Feris et al., 2004). In how far Acidobacteria may be involved in the degradation of pollutants under analysis here remains to be elucidated.

Figure 1: Community composition over time of (A) CZE soil samples, (B) BRA soil samples and (C) SUI soil samples, treated with benzene and BTEX, and of an untreated control. The relative abundance of sequence reads representing bacterial phyla/classes is given.

The high stability of the microbial community in the CZE soil was expected given that it has been under bioremediation since years (Junca & Pieper, 2004) and consequently the community in this soil should have been adapted to the degradation of aromatic compounds. In agreement, catabolic genes detected on the microarray are related to those observed in Proteobacteria (Figure 1A) and signals were observed in the benzene/toluene/isopropylbenzene dioxygenase encoding genes as well as those encoding enzymes of the EXDO-D branch of catechol 2,3-dioxygenases as previously described (Brennerova et al., 2009). In addition, signals of probes indicating the presence of proteobacterial catechol 1,2-dioxygenases and soluble diiron monooxygenases supported taxonomic results.

A comparison to the control microcosms incubated without additional pollutant suggests that the BTEX stress results in the selection of genes encoding benzene/toluene/isopropylbenzene dioxygenases related to YP556409 α-subunits of biphenyl dioxygenase from *Burkholderia xenovorans* LB400, Q46372 from *Comamonas testosteroni* B-356 and BAC01052 from *C. testosteronI* TK102 (Figure 2). However, after long-term contamination no signals could be recorded from this subfamily of genes, suggesting that other catabolic genes encoding enzymes capable of aromatic activation, which are not represented on the array have been selected. Supporting this hypothesis one probe (BAH90326) encoding a catechol 2,3-dioxygenase related to those of γ and β Proteobacteria including WP019451462 extradiol dioxygenase from *Cupriavidus* sp. BIS7 (Hong et al., 2012) showed a hybridization signal of 10 fold the intensity in the microcosm supplemented with benzene and of 12 fold the intensity in the microcosms supplemented with BTEX after long term contamination, compared to the control microcosm. This extradiol dioxygenase encoding gene in BIS7 is preceded by genes encoding a phenol monooxygenase which in turn are preceded by genes encoding a toluene monooxygenase in the same operon (WP019451463-75). Since this gene information was just recently published, probes covering the respective monooxygenases were not printed on the array. Thus, even though the taxonomical composition of the soil kept constant, the hybridization pattern suggests an adaptation of the catabolic route responsible for benzene or BTEX degradation.

Table 3: Microarray hybridization profiles of CZE soil samples in control, benzene- and BTEX-treated microcosms. The normalized signal intensity (NI) of probes showing hybridization after the indicated incubation time is given.

Family name		Control				Benzene					BTEX				
		Incubation days				Incubation days					Incubation days				
		2	4	10	20	2	4	20	30	90	4	10	20	60	90
Rieske non-heme iron oxygenase															
Accession number	AAB88813	0,4	0,6	0,4	0,8	0,3	1,0	1,0	0,5		1,2	0,8	3,0	1,0	0,3
	ACF20634		0,5	0,3	0,3	0,1	0,4	0,4	0,4						
	AAK14781		0,4				0,5	1,5	0,7		0,6	0,5	1,0		0,1
	YP556409					0,1	0,5	0,5	0,5			0,5	1,1		
	Q46372							0,6	0,8			0,1	0,8		
	BAC01052						0,1	0,4				0,1	0,4	1,2	
Extradiol dioxygenase															
Accession number	AAB88079	4,7	4,6	8,5	5,7	1,8	7,2	3,4		0,6	6,4	6,4	4,9	3,4	2,0
	YP001110005	1,9	1,7	1,9	1,8	0,7	4,6	1,4	0,2	0,2	3,7	1,5	2,4		0,7
	BAH90326		0,9		0,5			5,7	0,6	10,2	0,8	0,4			11,9
	BAH90119	2,2	3,8	4,0	11,5		0,9	28,2	5,1	3,0	2,8		21,8	5,7	1,2
Intradiol dioxygenase															
Accession number	YP587012	1,6	1,0	0,7	3,0	0,5	1,3	4,7		1,2	1,1	1,7	3,6	1,8	0,8
	YP001063703	0,8	0,6	0,4	1,4	0,1	0,7	3,0	2,0	1,1	0,6	0,9	2,7		0,7

Changes in community and catabolic gene structure during benzene and BTEX stress in BRA soil

In contrast to CZE soil, the bacterial community of the BRA soil reacted dramatically to incubation with benzene, and a clear increase in the relative abundance from 2 to 53%, of Actinobacteria was observed, which correlates with a decrease in the relative abundance from 68 to 22%, of β-Proteobacteria, specifically of members of the Comamonadaceae and Hydrogenophilaceae families. A similar trend was observed during BTEX incubation, where actinobacterial phylotypes were also enriched from 3 to 37% and β-proteobacterial phylotypes decreased from 61 to 35% (Figure 1-B).

A detailed analysis of phylotypes being enriched (Figure 3) during incubation with benzene showed six actinobacterial phylotypes which increased in relative abundance, four belonging to the Micrococcineae suborder and two to the Nocardiaceae family. Phy18 indicative for the presence of bacteria from the Micrococcaceae family was present in the original soil at a relative abundance as low as 0.02% but increased to a relative abundance of >20% after 90 days of incubation

with benzene. Members of the *Arthrobacter* genus inside the Micrococcaceae family have been described to degrade pollutants like methylpyridine (O'Loughlin et al., 1999) and benzene (Fahy et al., 2008). Phy30 indicative for the presence of *Rhodococcus spp*., not detectable in the original soil even at the high resolution level chosen, increased to a relative abundance of >10% of the total community. Various *Rhodococcus* strains are well documented for their outstanding capabilities to degrade aromatics, among them *Rhodococcus jostii* RHA1, a versatile biphenyl degrader (Seah et al., 2001), the 16S rDNA gene of which is identical over the sequence stretch analyzed to that of the mentioned *Rhodococcus* Phy30. Phy13 indicative for the presence of *Nocardia* spp. increased in abundance from 0,03% to roughly 16% and Phy4, which could only be classified to the suborder as Micrococcinae bacterium, increased in abundance from 1% to roughly 6% in 30 days and from then on diminished slightly.

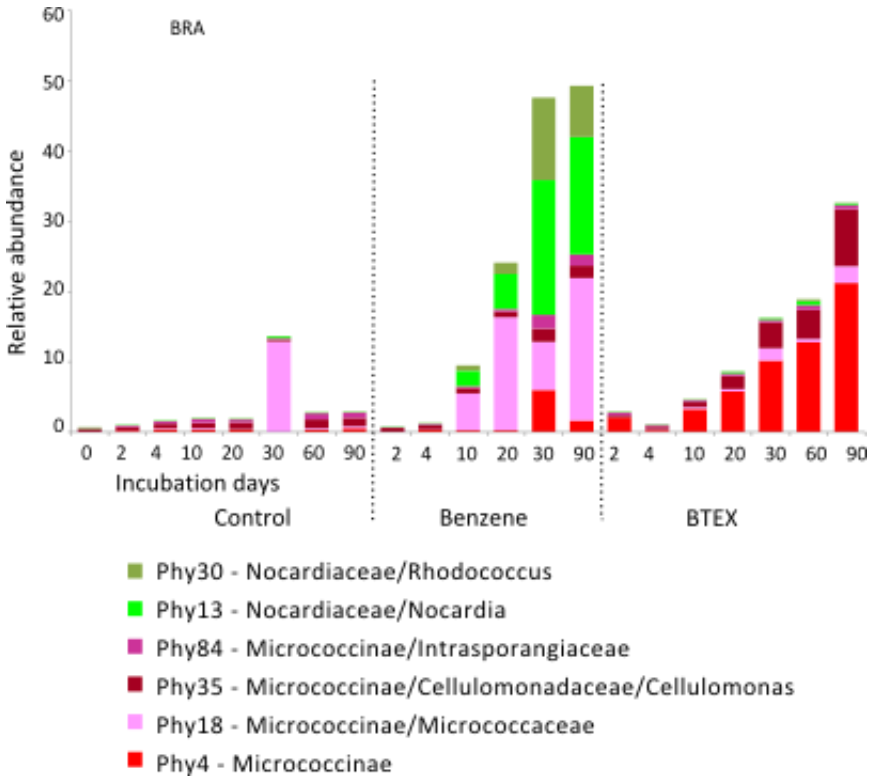


Figure 2: Community composition over time of BRA soil samples. The relative abundance of sequence reads of selected actinobacterial phylotypes is given.

Similarly, actinobacterial phylotypes were also enriched in microcosms treated with BTEX. Specifically, Phy4 mentioned above and Phy35, both indicative for the enrichment of organisms of the Micrococcineae suborder, increased constantly in relative abundance from 2 to 29% and from 0,2% to 8%, respectively. Phy35 (*Cellulomonas* sp.) was also enriched in the benzene microcosm but in lower abundance from 0,4% to 1,8%. *Cellulomonas* sp. has also been previously indicated as capable to degrade aromatic compounds and even benzene (Hernandez et al., 1997).

The enrichment of Actinobacteria during incubation of soil with benzene or BTEX was also reflected in the catabolic gene content (Figure 4). According to this, probes indicating the presence of genes encoding enzymes related to naphthalene dioxygenase of *Rhodococcus* sp. NCIMB 12038 (AAD28100) and catechol 2,3-dioxygenase (BAC00792) of the dibenzofuran degrader *Rhodococcus* sp. YK2 were enriched in both types of microcosms, those supplemented with benzene and those supplemented with BTEX.

Moreover, during extended incubation with benzene, genes encoding actinobacterial biphenyl or isopropylbenzene dioxygenases (related to *R. jostii* RHA1 YP707265 and *Rhodococcus erythropolis* BD2 NP898768) were obviously enriched suggesting that additional dioxygenases were recruited to attack benzene. Furthermore various catabolic gene probes indicating the presence of central metabolic routes of actinobacteria showed increased levels of hybridization in microcosms supplemented with benzene or BTEX (two probes, CAA67941 and AAT40306 indicating the enrichment of genes encoding catechol 1,2-dioxygenases related to those of *Rhodococcus opacus* 1CP and *R. erythropolis* SK121, respectively; probe NP601604 indicating the enrichment of genes encoding benzoate dioxygenases related to the enzyme of *Corynebacterium glutamicum*; probe AAX40648 indicating the enrichment of genes encoding hydroxyquinol 1,2-dioxygenases related to the enzyme of *Arthrobacter chlorophenolicus* A6 and probe ABA61848 indicating the enrichment of genes encoding maleylacetate reductases related to the enzyme of a *Pimelobacter simplex* strain 3E).

Table 4: Microarray hybridization profiles of BRA soil samples in control, benzene- and BTEX-treated microcosms. The normalized signal intensity (NI) of probes showing hybridization after the indicated incubation time is given.

Family name		Control					Benzene				BTEX		
		Incubation days					Incubation days				Incubation days		
		2	4	30	60	90	2	4	60	90	2	4	60
Rieske non-heme iron oxygenase													
Accession number	YP707265					0,1				0,8	0,1		
	NP898768							0,1		1,0			
	AAR05106		0,2			0,1	0,1			0,7	0,6	0,1	
	AAD28100		0,1	0,8	0,3	0,1			1,6	1,9			3,3
	BAB62286		0,1	2,7	0,7			0,2	3,3			1,0	1,5
	NP601604			0,4	2,2							0,1	15,5
	NP542871	0,3	0,1	5,5	1,3	0,2	0,2	0,4	1,4		0,1	0,9	10,0
	NP251208	0,7	0,8	3,3		0,4		2,8	0,7	0,7		3,2	10,9
	AAK52287	0,3		3,4	1,2	0,1		0,7				1,4	8,2
	YP260958	0,4	0,2	3,2	1,2	0,1	0,3	0,6				1,6	9,0
	YP348053		0,1	1,4		0,1		0,5				0,8	7,0
	YP348699		0,1	1,8	2,3	0,1		0,6				0,9	19,8
	AAX47023	0,2	0,1	4,3	4,1			0,5			0,1	3,3	16,7
	AAF63448	0,8	0,5	1,7	1,9	0,2	0,2	2,5	1,5	0,7	0,1	1,2	7,3
	NP745305	0,2		1,5	3,8		0,1	0,7	0,5			1,3	21,1
Ring hydroxylating monooxygenase													
Accession number	CAB55825	0,1	0,1	0,7	0,9					0,3		0,1	1,1
	BAA34172	0,1	0,1	0,8				0,2	2,1	0,6	0,1	0,1	5,1
	CAA55663		0,1	0,5					1,9				1,0
	YP001110001	0,2	0,1			1,7					0,1		
	AAG40794		0,1	0,2								0,1	0,6
Extradiol dioxygenase													
Accession number	BAH90119	3,4	6,0		4,2	2,1	1,1	3,5	3,8	1,2	1,4	5,8	5,4
	BAC00792		0,1	0,2	0,4			0,1	1,7	0,4		0,1	5,5
	AAA20982		0,1	0,5								0,2	6,4
	BAD72667	0,3	1,3					1,2	16,9	12,4		1,0	96,0
	BAD83302		0,6	4,0	2,9		0,3	0,8	9,2	5,6		1,0	35,4
	BAH89647		0,5	1,4	2,3	0,3		0,4	4,3		0,1		15,0
	ACH41239	0,1		1,7	8,7	0,2		0,5	9,7	4,2		0,5	46,4
	BAH90326					1,3		0,7		13,7	0,7	0,5	
Intradiol dioxygenase													
Accession number	CAA67941						0,1		1,9	1,5			3,7
	AAT40306		0,1	2,2		0,1		0,2	7,5			0,2	19,1
	AAX40648		0,1	0,9	1,8			0,4	6,0	1,7		0,1	11,7
	YP587012	0,3	0,7	3,0		0,6		1,0	3,7	4,9	0,1	1,2	18,0
	YP001063703	0,4	0,3	4,7	5,8	0,4		0,7	7,9	2,9	0,1	0,6	33,4
	CAF32822		0,7		0,6	0,2		0,9	4,5	3,5			16,7
	AAC35836		0,3	2,4				0,6	7,9	5,5			18,4
	AAR21663			1,3		0,1		0,4	15,7	2,6			19,0

Changes in community and catabolic gene structure during benzene and BTEX stress in SUI soil

The overall changes at the phylum/class level (Figure 1C) in SUI soil over time are as rapid as the reactions observed in the BRA microcosms. However, in contrast to BRA benzene treated microcosms, there was an extensive increase in the relative abundance of γ-proteobacteria during the first 10 days of the experiment, from 18 to 70%. Afterwards, similarly to what was observed in the BRA microcosms, actinobacterial phylotypes increased from 6 to 54% in the first 30 days, whereas β-proteobacterial phylotypes decreased from 45 to 17%.

In the benzene treated microcosms, an interesting succession of phylotypes could be observed (Figure 5). Phylotypes indicating the presence of *Pseudomonas* spp. were enriched during the first days of incubation and reached more than 70% of the total abundance in 10 days. Phy8, indicating the presence of *Pseudomonas* strains related to *P. pseudoalcaligenes* was the most abundant phylotype and accounted for 49 and 32% of sequence reads after 10 and 20 days of treatment, respectively. Members of the *Pseudomonas* genus are in fact widespread in BTEX contaminated sites (Stapleton et al., 2000; Junca et al., 2004; Witzig et al., 2006; Kabelitz et al. 2009) and different BTEX metabolic pathways have been described (Gibson et al., 1968; Williams and Murray, 1974; Yen et al., 1991).

Although the overall relative abundance of β-proteobacterial phylotypes decreased from 45 to 17% in benzene microcosms, a detailed analysis showed that 3 phylotypes were enriched during the incubation with benzene (Figure 5). These phylotypes indicate the presence of members of the Alcaligenaceae (Phy23), and Comamonadaceae (Phy11 and Phy32) families. Members of these families have previously been described as degraders of pollutants such as naphthalene, biphenyl or chlorobenzene or even as key players in the degradation at contaminated sites (Hiraoka et al., 2002; Yagi et al., 2009; Gross et al, 2008; Táncsics et al., 2011). As examples, *Hydrogenophaga*, a genus of the Comamonadaceae family, was identified by stable isotope probing as one of the phylotypes assimilating carbon from benzene in a contaminated site (Jechalke et al., 2013) and *Achromobacter* of the Alcaligenaceae family as a predominant biphenyl mineralizer (Sul et al., 2009)

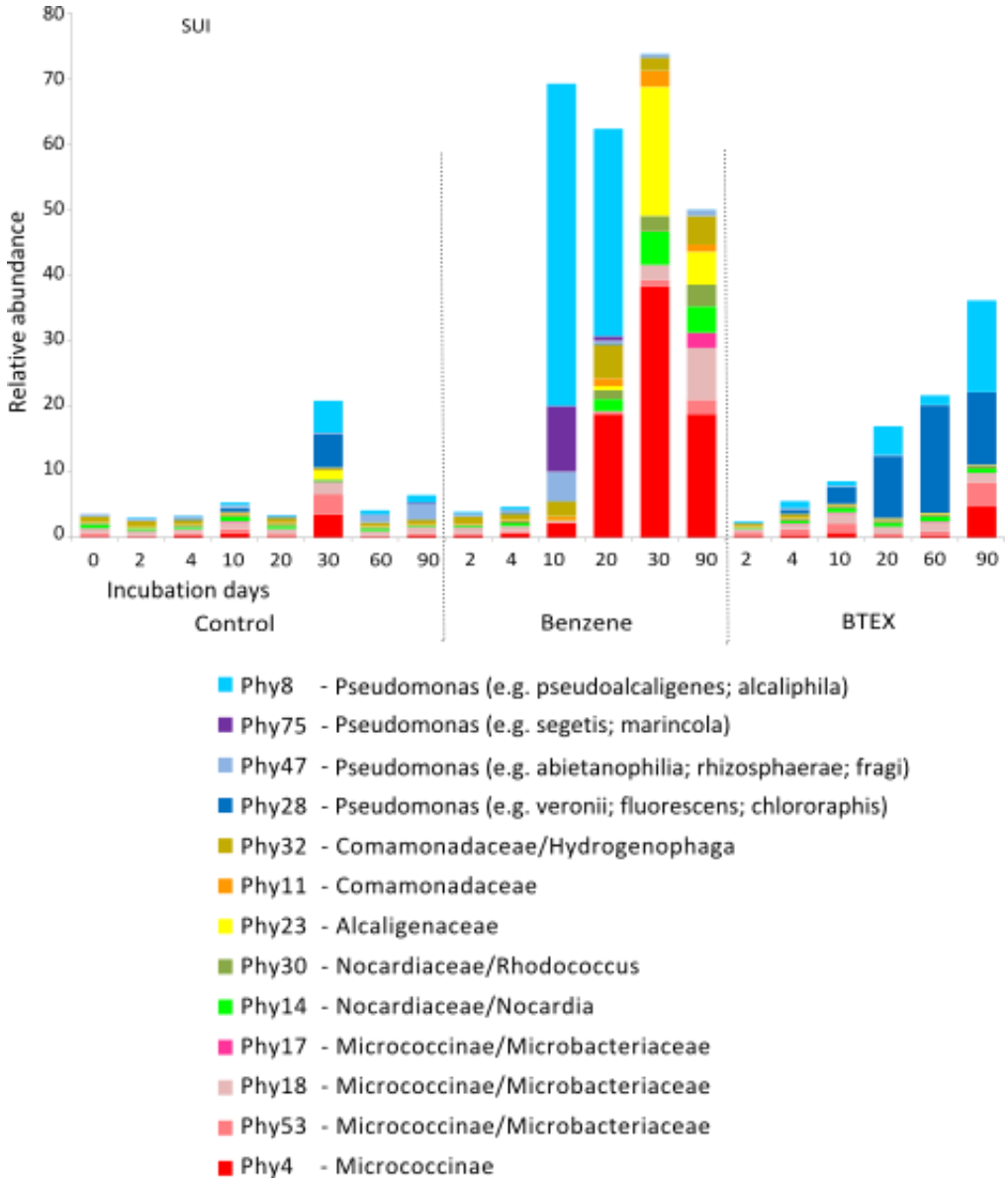


Figure 3: Community composition over time of SUI soil samples. The relative abundance of sequence reads of selected proteobacterial and actinobacterial phylotypes is given.

After 10-20 days, actinobacterial phylotypes started to increase in relative abundance. Even though some of those phylotypes were shared between BRA and SUI soils, such as a Phy30 and Phy13 indicating the presence of *Rhodococcus* sp. and *Nocardia* sp., respectively, the soil communities being enriched were clearly different. The most abundant actinobacterial phylotypes (Phy4), also being enriched in BRA microcosm treated with BTEX, belongs to the Micrococcinae suborder. It increased more than 100 times in 30 days (from 0,3% to 38%). Other phylotypes increasing in relative abundance also belong to the Micrococcinea suborder, but could be classified

as belonging to the Microbacteriaceae family and sum up to > 15% of relative abundance after 90 days of incubation with benzene.

Compared to the benzene treatment, the response of the SUI soil community to BTEX stress seems to be slower. Phy8 indicated the presence of microorganisms related to *P. pseudoalcaligenes*, increased in abundance over time in the BTEX treated microcosms and after extended incubation for 90 days accounted for 14% of the observed sequence reads. The same phylotype increased in relative abundance after 10 days in the microcosm incubated with benzene. A distinct phylotype (Phy28) indicating the presence of members of the *Pseudomonas fluorescens* group was observed to increase exclusively during BTEX treatment. These organisms increased in relative abundance from <0.1% to 16% in 60 days, and thereafter started to decline. Interestingly, *P. veronii* strains belonging to this group were described as abundant in a BTEX contaminated area previously (Witzig *et al*., 2006). Actinobacterial phylotypes started to increase only after extended incubation.

Analysis of the catabolic gene landscape also indicated *Pseudomonas* strains to become abundant after 10 days of incubation with benzene, as evidenced by the high abundance of signals indicating the presence of genes encoding proteins similar to catechol 1,2-dioxygenase of *Pseudomonas stutzeri* AN10 (CAE84865) or benzoate 1,2-dioxygenase of *Pseudomonas sp*. S-47 (AAO19113) (Figure 6). Concomitant with the obvious enrichment of Pseudomonas strains two probes indicating the presence of genes encoding catechol 2,3-dioxygenases with similarity to those of *P. stutzeri* AN10 (AAD02148) and *Pseudomonas sp*. IC (AAA20982) showed hybridization, as well as probes targeting genes encoding naphthalene dioxygenases related to those of *P. stutzeri* AN10 (AAD02136) and *Pseudomonas aeruginosa* PaK1 (Q51494), indicating *Pseudomonas* strains with the capability to degrade hydrophobic aromatic pollutants to become abundant.

Table 5: Microarray hybridization profiles of SUI soil samples in control, benzene- and BTEX-treated microcosms. The normalized signal intensity (NI) of probes showing hybridization after the indicated incubation time is given.

Family name		Control					Benzene		BTEX	
		Incubation days					Incub.		Incub.	
		2	4	10	20	90	20	90	2	90
Rieske non-heme iron oxygenase										
Accession number	AAD28100							0,8		
	AAD02136						2,1			
	Q51494						2,1			
	AAO19113		0,2				1,1			
	NP542871		0,4				1,0			
	YP709342		0,3				1,0			
	YP495819									1,7
Extradiol dioxygenase										
Accession number	BAA06872							0,7		
	AAA20982						4,3			
	AAD02148		0,3				4,3			
	BAD72667					0,6	6,5	5,2	1,9	
	BAH90326							7,8		
Intradiol dioxygenase										
Accession number	CAA67941							0,5		
	AAT40306		0,5					1,3		
	AAT40306b		0,4					1,0		
	AF003947		0,5					1,0		
	CAE84865						0,4			
	YP587012		1,0	3,6	0,9	0,3		2,7	1,0	
	YP001063703		0,8	2,4		0,3		1,7		
	CAF32822		1,6		1,5			1,4	1,2	
	AAC35836		0,5				0,4	2,6		

Similarly, the enrichment of actinobacterial phylotypes during long-term contamination with benzene is also underlined by the enrichment of actinobacterial catabolic encoding genes such as genes encoding catechol 1,2-dioxygenases similar to those of *R. opacus* 1CP (CAA67941) or *Nocardia* sp. H17-1 (AAT40306). However, besides such genes most probably fortuitously enriched, also genes encoding enzymes probably involved in hydrophobic pollutant degradation such as those encoding enzymes similar to 2,3-dihydroxybiphenyl dioxygenase of *R. jostii* RHA1 (BAA06872) or naphthalene 1,2-dioxygenase of *Rhodococcus* sp. NCIMB 12038 (AAD28100) were enriched during extended incubation with benzene, indicating actinobacteria to be active in degradation of added pollutants.

4.5 Conclusions

The reaction in CZE microcosms is slight when compared to BRA and SUI microcosm. Due to long term remediation (Junca & Pieper, 2004), the community is stable and mainly composed by β- and α-Proteobacteria, however, evidence for an adaptation of the metabolic web to the novel pollutant mixtures could be given. BRA and SUI soil harboured similar bacterial communities, and in both soils extended incubation resulted in an enrichment of Actinobacteria, which was also reflected in the enrichment of genes supposedly characterizing the core genome, but also specific oxygenases dedicated to the metabolism of hydrophobic pollutants.

However, whereas in BRA microcosms, actinobacteria obviously directly took over the lead in biodegradation, a succession of bacterial communities was clearly observed in SUI microcosms where initially γ-proteobacterial phylotypes, related to *Pseudomonas* spp. were enriched which after extended incubation were substituted by β-proteobacterial and actinobacterial phylotypes. Even though different phylotypes were enriched in BRA and SUI microcosms, there were also important similarities in the enriched phylotypes and catabolic genes, probably indicative for their environmental fitness.

Clearly different responses of the soil communities to stress exerted by the incubation with benzene and BTEX were observed. This may reflect their different toxicity, but also their different metabolism despite similar structures. Interestingly, the response to benzene stress was generally faster compared to the response to BTEX stress.

4.6 Acknowledgements

This work was funded by BACSIN (project number: 211684) from the European Commission. Daiana de Lima Morales was supported by a grant from the CNPq-Conselho Nacional de Desenvolvimento Científico e Tecnológico-Brazil (project number: 290020/2008-5). Special thanks to Dr. Robert Geffers from the sequencing facility at the Helmholtz Centre for Infection Research and to Agnes Waliczek for technical assistance.

4.7 References

Anderson I, Held B, Lapidus A, Nolan M, Lucas S, Tice H, Glavina del Rio T, Cheng J-F, Han C, Tapia R, Goodwin L, Pitluck S, Liolios K, Mavromatis K, Pagani I, Ivanova N, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Brambilla E, Rohde M, Spring S, Göker M, Detter JC, Woyke T, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Klenk H-P & Kyrpides NC (2012) Genome sequence of the homoacetogenic bacterium *Holophaga foetida* type strain (TMBS4(T)). *Stand Genomic Sci* **6**: 174–184.

Berthrong S, Schadt C W, Piñeiro G & Jackson RB (2009) Afforestation alters the composition of functional genes in soil and biogeochemical processes in South American grasslands. *Appl Environ Microbiol* **75**: 6240–6248.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA & Caporaso JG (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* **10**: 57–59.

Brennerova MV, Josefiova J, Brenner V, Pieper DH & Junca H (2009) Metagenomics reveals diversity and abundance of *meta*-cleavage pathways in microbial communities from soil highly contaminated with jet fuel under air-sparging bioremediation. *Environ Microbiol* **11**: 2216–2227.

Camarinha-Silva A, Wos-Oxley ML, Jáuregui R, Becker K & Pieper DH (2012) Validating T-RFLP as a sensitive and high-throughput approach to assess bacterial diversity patterns in human anterior nares. *FEMS Microbiol Ecol* **79**: 98–108.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G & Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**: 1621–1624.

Corseuil HX, Dal M & Marins M (1997) Contaminação de águas subterrâneas por derramamentos de gasolina: o problema é grave? *Revista Engenharia Sanitária e Ambiental* **2**: 50–54.

Degnan PH & Ochman H (2012) Illumina-based analysis of microbial community diversity. *ISME J* **6**: 183–194.

Eaton RW, Selifonova OV & Gedney RM (1998) Isopropylbenzene catabolic pathway in *Pseudomonas putida* RE204: nucleotide sequence analysis of the *ipb* operon and neighboring DNA from pRE4. *Biodegradation* **9**: 119–132.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Fahy A, Ball AS, Lethbridge G, McGenity TJ & Timmis KN (2008) High benzene concentrations can favour Gram-positive bacteria in groundwaters from a contaminated aquifer. *FEMS Microbiol Ecol* **65**: 526–533.

Feris KP, Hristova K, Gebreyesus B, Mackay D & Scow KM (2004) A shallow BTEX and MTBE contaminated aquifer supports a diverse microbial community. *Microbial Ecol* **48**: 589–600.

Gibson DT, Koch JR & Kallio RE (1968a) Oxidative degradation of aromatic hydrocarbons by microorganisms. I. Enzymic formation of catechol from benzene. *Biochemistry* **7**: 2653–2662.

Gibson DT, Koch JR, Schuld CL & Kallio RE (1968b) Oxidative degradation of aromatic hydrocarbons by microorganisms. II. Metabolism of halogenated aromatic hydrocarbons. *Biochemistry* **7**: 3795–3802.

Greated A, Lambertsen L, Williams PA & Thomas CM (2002) Complete sequence of the IncP-9 TOL plasmid pWW0 from P*seudomonas putida*. *Environ Microbiol* **4**: 856–871.

Gross R, Guzman CA, Sebaihia M, dos Santos VA PM, Pieper DH, Koebnik R, Lechner M, Bartels D**,** Buhrmester J, Choudhuri JV, Ebensen T, Gaigalat L, Herrmann S, Khachane AN, Larisch C, Link S, Link, B, Meyer F, Mormann S, Nakunst D, Rückert C, Schneiker-Bekel S, Schulze K, Vorhölter F-J, Yevsa T, Engle JT, Goldman WE, Pühler A, Göbel UB, Goesmann A, Blöcker H, Kaiser O & Martinez-Arias R (2008) The missing link: *Bordetella petrii* is endowed with both the metabolic versatility of environmental bacteria and virulence traits of pathogenic Bordetellae. *BMC Genomics* **9**: 449.

Hamady M, Walker JJ, Harris JK, Gold NJ & Knight R (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Meth* **5**: 235–237.

He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL, Li X, Wu L, Gentry TJ, Yin Y, Liebich J, Hazen TC & Zhou J (2010) GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J* **4**: 1167–1179.

He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC, Huang Z, Wu W, Gu B, Jardine P, Criddle C & Zhou J (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1**: 67–77.

Hendrickx B, Junca H, Vosahlova J, Lindner A, Rüegg I, Bucheli-Witschel M, Faber F, Egli T, Mau M, Schlömann M, Brennerova M, Brenner V, Pieper DH, Top EM, Dejonghe W, Bastiaens L & Springael D (2006) Alternative primer sets for PCR detection of genotypes involved in bacterial aerobic BTEX degradation: distribution of the genes in BTEX degrading isolates and in subsurface soils of a BTEX contaminated industrial site. *J Microbiol Meth* **64**: 250–265.

Hernandez BS, Koh S-C, Chial M & Focht DD (1997) Terpene-utilizing isolates and their relevance to enhanced biotransformation of polychlorinated biphenyls in soil. *Biodegradation* **8**: 153–158.

Hiraoka Y, Yamada T, Tone K, Futaesaku Y & Kimbara K (2002) Flow cytometry analysis of changes in the DNA content of the polychlorinated biphenyl degrader *Comamonas*

testosteroni TK102 : Effect of metabolites on cell-cell separation. *Appl Environ Microbiol* **68**: 5104-5112.

Hong KW, Thinagaran DAl, Gan HM, Yin W-F & Chan KG (2012) Whole-genome sequence of *Cupriavidus* sp. strain BIS7, a heavy-metal-resistant bacterium. *J Bacteriol* **194**: 6324.

Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA Genes. *Appl Environ Microbiol* **72**: 1719–1728.

Jechalke S, Franchini AG, Bastida F, Bombach P, Rosell M, Seifert J, von Bergen M, Vogt C & Richnow HH (2013) Analysis of structure, function, and activity of a benzene-degrading microbial community. *FEMS Microbiol Ecol* **85**: 14–26.

Jeon C O, Park W, Padmanabhan P, DeRito C, Snape JR & Madsen EL (2003) Discovery of a bacterium, with distinctive dioxygenase, that is responsible for in situ biodegradation in contaminated sediment. *Proc Natl Acad Sci U S A* **100**: 13591–13596.

Junca H & Pieper DH (2010) Functional marker gene assays for hydrocarbon degrading microbial communities: Aerobic. In *Handbook of Hydrocarbon and Lipid Microbiology*. Timmis, KN (ed.). Berlin Heidelberg, Germany: Springer-Verlag, 4289-4312.

Junca H, Plumeier I, Hecht H-J & Pieper DH (2004) Difference in kinetic behaviour of catechol 2,3-dioxygenase variants from a polluted environment. *Microbiology* **150**: 4181–4187.

Kabelitz N, Machackova J, Imfeld G, Brennerova M, Pieper DH, Heipieper HJ & Junca H (2009) Enhancement of the microbial community biomass and diversity during air sparging bioremediation of a soil highly contaminated with kerosene and BTEX. *Appl Microbiol Biotechnol* **82**: 565–577.

Keith L & Telliard W (1979) ES&T Special Report: Priority pollutants: I-a perspective view. *Environ Sci Technol* **13**: 416–423.

Kircher M, & Kelso J (2010) High-throughput DNA sequencing concepts and limitations. *Bioessays* **32**: 524–536.

Yuting Liang, Joy D Van Nostrand, Ye Deng, Zhili He, Liyou Wu, Xu Zhang, Guanghe Li & Jizhong Zhou (2011) Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *ISME J* **5**: 403–413.

Liesack W, Bak F, Kreft J-U & Stackebrandt E (1994) *Holophaga foetida* gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. *Arch Microbiol* **162**: 85–90.

Liu Z & Liu J (2013) Evaluating bacterial community structures in oil collected from the sea surface and sediment in the northern Gulf of Mexico after the Deepwater Horizon oil spill. *Microbiologyopen* **2**: 492–504.

Mason OU, Hazen TC, Borglin S, Chain PS, Dubinsky EA, Fortney JL, Han J, Holman HY, Hultman J, Lamendella R, Mackelprang R, Malfatti S, Tom LM, Tringe SG, Woyke T, Zhou J, Rubin EM & Jansson JK (2012) Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *ISME J* **6**: 1715–1727.

Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M & Daniel R (2011) Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PloS One* **6**: e17000.

Naether A, Foesel BU, Naegele V, Wüst PK, Weinert J, Bonkowski M, Alt F, Oelmann Y, Polle A, Lohaus G, Gockel S, Hemp A, Kalko EK, Linsenmair KE,Pfeiffer S, Renner S, Schöning I, Weisser WW, Wells K, Fischer M, Overmann J & Friedrich MW (2012) Environmental factors affect Acidobacterial communities below the subgroup level in grassland and forest soils. *Appl Environ Microbiol* **78**: 7398–7406.

Nebe J, Baldwin BR, Kassab RL, Nies L & Nakatsu CH (2009) Quantification of aromatic oxygenase genes to evaluate enhanced bioremediation by oxygen releasing materials at a gasoline-contaminated site. *Environ Sci Technol* **43**: 2029–2034.

Newman LM & Wackett LP (1995) Purification and characterization of toluene 2-monooxygenase from *Burkholderia cepacia* G4. *Biochemistry* **34**: 14066–14076.

O'Loughlin E, Sims G & Traina S (1999) Biodegradation of 2-methyl, 2-ethyl, and 2-hydroxypyridine by an *Arthrobacter* sp. isolated from subsurface sediment. *Biodegradation* **10**: 93–104.

Pérez-Pantoja D, De la Iglesia R, Pieper DH & González B (2008) Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. *FEMS Microbiol Rev* **32**: 736–794.

Perez-Pantoja D, Donoso R, Junca H, Gonzalez B & Pieper DH (2009) Phylogenomics of aerobic bacterial degradation of aromatics. In *Handbook of Hydrocarbon and Lipid Microbiology*. Timmis, KN (ed.). Berlin Heidelberg, Germany: Springer-Verlag, 1356-1397.

Martin-Laurent F, Philippot L, Hallet S, Chaussod R, Germon JC, Soulas G & Catroux G (2001) DNA extraction from soils : Old bias for new microbial diversity analysis methods. *Appl Environ Microbiol* **67**: 2354–2359.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG,Van Horn DJ & Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Seah SYK, Labbé G, Kaschabek SR, Reifenrath F, Reineke W & Eltis LD (2001) Comparative specificities of two evolutionarily divergent hydrolases involved in microbial degradation of polychlorinated biphenyls. *J Bacteriol* **183**: 1511-1516.

Shim H, Shin E & Yang S-T (2002) A continuous fibrous-bed bioreactor for BTEX biodegradation by a co-culture of *Pseudomonas putida* and *Pseudomonas fluorescens*. *Adv Environ Res* **7**: 203–216.

Singleton DR, Richardson SD & Aitken MD (2011) Pyrosequence analysis of bacterial communities in aerobic bioreactors treating polycyclic aromatic hydrocarbon-contaminated soil. *Biodegradation* **22**: 1061–1073.

Soergel DaW, Dey N, Knight R & Brenner SE (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J* **6**: 1440–1444.

Stapleton RD, Bright NG & Sayler GS (2000) Catabolic and genetic diversity of degradative bacteria from fuel-hydrocarbon contaminated aquifers. *Microbial Ecol* **39**: 211–221.

Sul WJ, Park J, Quensen JF, Rodrigues JLM, Seliger L, Tsoi TV, Zylstra GJ & Tiedje JM (2009) DNA-stable isotope probing integrated with metagenomics for retrieval of biphenyl dioxygenase genes from polychlorinated biphenyl-contaminated river sediment. *Appl Environ Microbiol* **75**: 5501–5506.

Sutton NB, Maphosa F, Morillo JA, Abu Al-Soud W, Langenhoff AA, Grotenhuis T, Rijnaarts HH & Smidt H (2013) Impact of long-term diesel contamination on soil microbial community structure. *Appl Environ Microbiol* **79**: 619–630.

Vilchez-Vargas R, Geffers R, Suárez-Diez M, Conte I, Waliczek A, Kaser VS, Kralova M, Junca H & Pieper DH (2013) Analysis of the microbial gene landscape and transcriptome for aromatic pollutants and alkane degradation using a novel internally calibrated microarray system. *Environ Microbiol* **15**: 1016–1039.

Wang F, Zhou H, Meng J, Peng X, Jiang L, Sun P, Zhang C, Van Nostrand JD, Deng Y, He Z, Wu L, Zhou J & Xiao X (2009) GeoChip-based analysis of metabolic diversity of microbial communities at the Juan de Fuca Ridge hydrothermal vent. *Proc Natl Acad Sci USA* **106**: 4840–4845.

Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73***:* 5261–5267.

Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD, Bradley B, Brettin TS, Brinkac LM,Bruce D, Creasy T, Daugherty SC, Davidsen TM, DeBoy RT, Detter JC, Dodson RJ, Durkin AS, Ganapathy A, Gwinn-Giglio M, Han CS, Khouri H, Kiss H, Kothari SP, Madupu R, Nelson KE, Nelson WC, Paulsen I, Penn K, Ren Q, Rosovitz MJ, Selengut JD, Shrivastava S, Sullivan SA, Tapia R, Thompson LS, Watkins KL,Yang Q, Yu C, Zafar N, Zhou L & Kuske CR (2009)

Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microb* **75**: 2046–2056.

Werner JJ, Zhou D, Caporaso JG, Knight R & Angenent LT (2012) Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *ISME J* **6**: 1273–1276.

Whited GM & Gibson DT (1991) Separation and partial characterization of the enzymes of the toluene-4-monooxygenase catabolic pathway in *Pseudomonas mendocina* KR1. *J Bacteriol* **173**: 3017–3020.

Williams PA & Murray K (1974) Metabolism of benzoate and the methylbenzoates by *Pseudomonas putida* (arvilla) mt-2: evidence for the existence of a TOL plasmid. *J Bacteriol* **120**: 416–423.

Witzig R, Junca H, Hecht HJ & Pieper DH (2006) Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: Links between benzene degradation and genes similar to those encoding isopropylbenzene dioxygenases. *Appl Environ Microbiol* **72**: 3504–3514.

Yagi JM, Sims D, Brettin T, Bruce D & Madsen EL (2009) The genome of *Polaromonas naphthalenivorans* strain CJ2, isolated from coal tar-contaminated sediment, reveals physiological and metabolic versatility and evolution through extensive horizontal gene transfer. *Environ Microbiol* **11**: 2253–2270.

Yen KM, Karl MR, Blatt LM, Simon MJ, Winter RB, Fausset PR, Lu HS, Harcourt AA & Chen KK (1991) Cloning and characterization of a *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. *J Bacteriol* **173**: 5315–5327.

Chapter V - General Discussion

Benzene, toluene, ethylbenzene and xylene isomers, known as BTEX, are among the major aromatic hydrocarbons in petroleum products like gasoline, diesel fuel, lubricating and heating oil. Continuing exposure to high concentrations of BTEX can lead to damage of the liver, kidneys, central nervous system and eyes, in addition benzene is classified as a "Group A, known human carcinogen of medium carcinogenic hazard" by the US EPA (http://www.epa.gov/teach/chem_summ/BENZ_summary.pdf). BTEX compounds are frequently accidentally discharged into soil as a result of leakage through underground storage tanks and pipelines, improper waste disposal practices, inadvertent spills and leaching from landfills (Corseuil et al., 1997; Shim et al., 2002), (http://www.gsi-net.com/files/papers/meamodel.pdf) posing massive environmental threats. Consequently BTEX (together with chlorinated hydrocarbons), are the most relevant groundwater contaminants, with significant impact on drinking water according to The World Health Organization (WHO).

Using bacteria for treating contaminated sites is an energy-efficient and environmentally sound approach (Lapertot et al., 2000). However, despite increasing information, knowledge on in situ processes is still fragmentary. With the goal to better understand BTEX biodegradation in this work, I could; (1) genome sequence two key players (*Pseudomonas veronii* 1YdBTEX2 and *Pseudomonas veronii* 1YB2) known to be important for aromatic biodegradation, characterize the catabolic pathway in both strains and identify genetic determinants possibly important for the environmental success of these strains; (2) characterize the microbial community structure and structural changes of soils under aromatic pollutant stress and (3) elucidate the catabolic potential and behavior of soil communities under contaminant stress over time.

The reasons why some microorganisms successfully establish themselves during bioremediation processes is still to be elucidated. Here I drafted the genomes of *Pseudomonas veronii* 1YdBTEX2 and *Pseudomonas veronii* 1YB2, both isolated from soil, in the Czech Republic, highly contaminated with benzene (Junca & Pieper, 2003), purported to be key players in the biodegradation of pollutants at that site (Junca & Pieper, 2003; Witzig et al., 2006) Curiously, both strains harbor unique catabolic

pathways for the degradation of benzene, and only 1YdBTEX2 is able to degrade toluene. According to former studies, the difference in substrate specificity is due to small sequence differences in the alpha-subunit of the initial multicomponent Rieske non-heme iron oxygenase responsible for activation of the aromatic ring (Junca & Pieper, 2003; Witzig et al., 2006). Interestingly, this oxygenase is most closely related to isopropylbenzene dioxygenases and not to the benzene/toluene dioxygenases (Witzig et al., 2006). Another intriguing characteristic (Witzig et al. 2006) is that the catechol 2,3-dioxygenase, expressed by 1YB2 and 1YdBTEX2 during growth on benzene, belongs to subfamily I.2.A (EXDO-A) whereas an enzyme of the subfamily I.3.A (EXDO-K2), typically encoded in the same operon as an isopropylbenzene dioxygenase (or even a toluene dioxygenase) (Aoki et al., 1996; Eaton, 1996) was not induced.

According to our knowledge this is the first time the genome from a *Pseudomonas veronii* species was sequenced. With this data in hand I could accomplish the first aim of this thesis, the comparison of the genomes of *P. veronii* 1YdBTEX2 and *P. veronii* 1YB2. This allowed also, the elucidation of the organization of the benzene and/or toluene catabolic pathways the verification of the key genes to gain an overview on the metabolic properties and catabolic gene landscape of these key players for bioremediation and genome comparisons revealed why these isolates may prevail in an environment under pollutant stress.

According to the 16S rRNA sequence, *P. veronii* 1YdBTEX2 and 1YB2 are more similar to the type strain *P. veronii* CIP104663^T (99.67% and 99.74%) than to *P. fluorescens* IAM12022^T (98.7% and 98.8%). The same is found with the *gyr*B gene sequence, where *P. veronii* CIP104663^T is 99.4 % similar to both strains and *P. fluorescens* IAM12022^T is 91.0% similar to both strains. The data aforementioned corroborate with the fact that 1YdBTEX2 and 1YB2 should be considered as representatives of the *P. veronii* rather than of the *P. fluorescens* species.

In agreement with this result, the proteome analysis of the strains sequenced in this work (1YdBTEX2 and 1YB2) indicated that they are rather distantly related to *P. fluorescens* SBW25 and A506, which may represent “true” *P. fluorescens* strains. In agreement *P. fluorescens* A506 and SBW25 were grouped together before in a genome

multilocus sequence analysis (MLSA) (Loper et al., 2012). Other studies highlighted the need for taxonomic revision within the *P. fluorescens* group and emphasized that they are only distantly related and thus might constitute different species (Silby et al., 2009; Silby et al., 2011; Loper et al., 2012). Accordingly, *P. fluorescens* Pf-5 and *P. fluorescens* CHA0, were recently classified as *P. protegens* Pf-5 and CHA0 (Ramette et al., 2011) and grouped together in our proteome analysis. As well, our study suggests that isolates classified as *P. fluorescens* strains might be re-classified into other species. Indeed our proteome studies suggest that *P. fluorescens* F113 and *P. brassicacearum* NFM4 belong to the same species.

After comparing the *P. veronii* 1YdBTEX2 and 1YB2 genomes with other *P. fluorescens* genomes, the next goal was to elucidate the organization of the benzene and/or toluene catabolic pathways and to verify expression levels of key genes. From previous studies, part of the isopropylbenzene oxygenases genes (Ipb) (Witzig et al., 2006) and the *meta*-cleavage gene EXDO-A (subfamily I.2.A) had already been identified (Junca, et al., 2004). Albeit sequence information showed a complete *ipb* operon in both strains, encoding a Rieske non-heme iron oxygenase, a *cis*-isopropylbenzenedihydrodiol dehydrogenase and an EXDO-K2 type extradiol dioxygenase (subfamily I.3.A) (Eltis & Bolin, 1996), detailed analysis revealed a stop codon in the extradiol dioxygenase encoding gene of 1YB2, resulting in a non-functional gene product. Similarly a stop codon was observed in the 1YdBTEX2 gene encoding *cis*-isopropylbenzenedihydrodiol dehydrogenase, obviously abolishing also EXDO-K2 activity.

The inactivation of a gene encoding an extradiol dioxygenase in catabolic gene clusters encoding enzymes for the activation of hydrophobic aromatic pollutants has previously been described. As an example, the degradation of 1,2,4-trichlorobenzene by *Ralstonia* sp. PS12 is initiated by a Rieske non-heme iron oxygenase with high similarity to isopropylbenzene dioxygenases (Beil et al., 1997). However, the further degradation of central chlorocatechols is achieved by an intradiol cleavage chlorocatechol pathway, whereas extradiol dioxygenases are typically inactivated during chlorocatechol turnover (Bartels et al., 1984). To avoid the energy costing

synthesis of an extradiol dioxygenase, the respective gene is inactivated in strain PS12 (Beil et al., 1999).

The degradation of benzene via an extradiol cleavage pathway necessitates a 2-hydroxymuconic semialdehyde dehydrogenase (Harayama et al., 1987), however, respective genes are absent from toluene/isopropylbenzene/biphenyl catabolic gene clusters and only a 2-hydroxymuconic semialdehyde hydrolase is encoded there. Thus, a 2-hydroxymuconic semialdehyde dehydrogenase has to be recruited from elsewhere. In *P. veronii* 1YdBTEX2 a 2-hydroxymuconic semialdehyde dehydrogenase encoding gene was localized in a salicylate catabolic cluster together with a gene encoding a EXDO A type extradiol dioxygenase. In *P. veronii* 1YB2 EXDO A and 2-hydroxymuconic semialdehyde dehydrogenase encoding genes were localized in a truncated naphthalene gene cluster, where the naphthalene dihydrodiol dehydrogenase (*nahB*) is interrupted by a transposase and the initial dioxygenase is incomplete. Thus, recruitment of a gene cluster encoding an extradiol dioxygenase and a 2-hydroxymuconic semialdehyde dehydrogenase would be sufficient to allow growth on benzene and makes expression of the extradiol dioxygenase encoded in the isopropylbenzene dioxygenase gene cluster dispensable. However, the reason why also the cis-benzene dihydrodiol dehydrogenase encoded by this gene cluster is inactivated and why such an activity is recruited from elsewhere in both strains is not so obvious. Interestingly, in *Pseudomonas putida* ML2, one of the few bacterial strains isolated based on its capability to grow on benzene (Tan & Mason, 1990) and characterized in detail regarding enzymes involved in benzene degradation, the dehydrogenation of *cis*-benzene dihydrodiol is catalyzed by BedD, a *cis*-benzene dihydrodiol dehydrogenase with high similarity to glycerol dehydrogenases (Fong et al., 1996). This enzyme is encoded by genes located directly upstream of genes encoding the benzene dioxygenase whereas genes encoding enzymes related to toluene *cis*-dihydrodiol dehydrogenases of the short chain dehydrogenase family were absent. Curiously, degradation of intermediate catechol proceeds via an intradiol cleavage pathway (Tan & Fong, 1993) which belongs to the core genome of most *Pseudomonas* strains (Perez Pantoja et al., 2009). Similarly to the situation in strain ML2, dehydrogenation of intermediate *cis*-benzene dihydrodiol in 1YB2 and 1YdBTEX2

is catalyzed by enzymes which are not closely related to toluene *cis*-dihydrodiol dehydrogenases. Thus, in both *P. putida* ML2 as well as *P. veronii* 1YB2 and 1YdBTEX2 degradation of benzene is, except for the initial dioxygenation step, not catalyzed by enzymes similar to those involve in toluene degradation via a dioxygenolytic pathway. If this is due to poor activity of enzymes of the toluene dioxygenase pathway remains to be elucidated.

Besides the unique benzene degradation pathway, *P. veronii* 1YdBTEX2 and 1YB2 also harbour genetic attributes that may facilitate growth under unfavourable conditions such as a secondary siderophore synthesis gene cluster, a complete denitrification pathway organized in one superoperonic cluster and a gene cluster for the use of hydrogen as electron donor.

The synthesis of a second siderophore with a simple structure, besides pioverdine, might help *P. veronii* 1YB2 and 1YdBTEX2 to prevail in an iron limiting environment. Both strains harbour the genes for pyochelin synthesis, which show an organization identical to that found in *P. aeruginosa* PAO1 (ATCC 15692) with amino acid sequence identities of 40 % to 75%. Another advantage would be the capability to grow under anaerobic conditions. *P. veronii* 1YB2 and 1YdBTEX2 harbour all the genes necessary for complete denitrification. A partial denitrification pathway has being described in many *Pseudomonas* species, however, the capability to completely reduce nitrate to molecular nitrogen is a feature observed in only a few *Pseudomonas* species. Out of the 44 completely sequenced *Pseudomonas* spp (http://www.pseudomonas.com/), 17 harbour the complete denitrification pathway (NO_3^- to N_2), among them 9 *P. aeruginosa*, all 6 *P. stutzeri* strains, in addition to *P. brassicacearum* NFM421 and *P. fluorescence* F113. Interestingly, both strains are closely related, as indicated by the similarity of the gyrB gene sequence as well as the 16S rDNA sequence and including the proteomic analysis, indicating that they may belong to the same species. Denitrification was absent from all 11 sequenced *P. putida* strains, both *P. veronii* strains analyzed here show a gene organization similar to the one observed in *P. fluorescens* F113 (Redondo-Nieto et al., 2012) and highest sequence identity of the encoded proteins to those of *Pseudomonas extremaustralis* 14-3 (Tribelli et al., 2012) where sequence identity was >94%, though *P. extremaustralis* 14-

3 does not harbour a complete denitrification cluster, as the nitrite reductase gene cluster is absent. Importantly in 1YdBTEX2 and 1YB2, all genes clusters (Nrd, Nar, Nos, Nor and Nir) involved in denitrification reactions, are localized in one superoperonic gene cluster.

A further exceptional feature would be to obtain energy from the oxidation of H_2. In fact, *P. veronii* 1YB2 and 1YdBTEX2 harbour a membrane bound [NiFe]-hydrogenase gene cluster comprising genes encoding the structural enzymes (*hoxKGZ*), accessory proteins (*hoxMLOQRTV*) responsible for Ni-Fe center assembly (*hypABFCDEX*), a regulatory system (*hoxABCJ*) and a high affinity nickel transporter (*hoxN1*). This indicates that both *P. veronii* strains may use hydrogen as electron donor. Moreover, a denitrification capability can be linked to the use of H_2 such as the case in *Cupriavidus necator* H16. If *P. veronii* strains isolated here may perform such reaction remains to be elucidated. Interestingly, it is hypothesized that *Methylomirabilis oxyfera* is capable of producing oxygen via a new intra-aerobic pathway that involves the conversion of two nitric oxide molecules to dinitrogen and molecular oxygen (Haroon et al., 2013). The intra-cellular molecular oxygen produced, is used to oxidize methane via the well studied aerobic methane monooxygenases (Ettwig et al., 2010). Taking into consideration these studies on *Methylomirabilis oxyfera* it can be postulated that an “anaerobic pathway” using intra-cellular oxygen, for benzene degradation, might exist. Such a pathway using “intracellular oxygen” has been described for benzene degradation by *Alicycliphilus denitrificans* strains BC and K601 (Weelink et al., 2008; Oosterkamp et al., 2013), where oxygen formed by dismutation of chlorite is used to oxygenate benzene. If *P. veronii* strains isolated here may perform denitrification with hydrogen as electron donor and anaerobic benzene degradation remains to be elucidated.

Whereas the study of single microorganisms gives detailed information on isolates assumed as important in bioremediation processes, the monoculture of bacteria using contaminants as sole carbon source does not describe the in vivo community dynamics. Thus, the final aim of this thesis was to analyze the microbial community and catabolic gene structures of contaminated soils of different origins under constant and specific contamination with benzene and BTEX in microcosm

experiments over time and to identify shifts in community structure and prevalence of key genes.

Eliminating the need of cultivation, molecular techniques allow studying the microbial diversity and activity in situ. Many molecular fingerprinting techniques (e.g. DGGE, SSCP), however, are laborious and when targeting functional genes, they are restricted by the use of primers for amplification, which cover only a small part of the microbial diversity. The development of functional gene arrays solve the limitations of former fingerprinting techniques by permitting the hybridization based screening of hundreds of different genes at once. The microarray, developed by our group, is a powerful genomic technology that generates massive information from environmental samples (Vilchez-Vargas et al., 2013). The array is based on upgraded databases of key aromatic catabolic genes (Perez-Pantoja et al., 2009), avoiding the huge level of misannotations found in databases (Vilchez-Vargas et al., 2013), which results in misleading probes and incorrectly interpreted results, such as the case for the GeoChip (Berthrong et al.,2009; He et al., 2010).

A next generation amplicon sequencing protocol has recently been developed by our group, which allows a high-throughput cost effective microbial community analysis (Camarinha-Silva et al., 2013). To analyze which microorganisms were important in the BTEX contaminated soil and selected under contaminant stress, we applied the developed protocol, however using the V5/V6 hypervariable region of the 16S rRNA gene. Based on an in silico comparison against the entire RDP database, the primer set amplifying the V5/V6 region could anneal and amplify 99.96% of bacterial and 97.82% of archaeal sequences, allowing one mismatch (Bohorquez et al., 2012). With above described techniques at our reach, we designed a microcosms experiment, where three contaminated soil samples, from Czech Republic (CZE), Brazil (BRA) and Switzerland (SUI), were incubated during 90 days, under constant benzene and BTEX stress.

Since CZE microcosms has been under bioremediation since decades (Junca & Pieper, 2004) the adapted community composed mainly of β- and α-Proteobacteria was maintained during all the experiment and only one phylotype related to Acidobacteria gradually increased in relative abundance in benzene treated

microcosms. It is important to point out that due to the difficulty to culture Acidobacteria strains, little is known about members of this phylum and genes involved in aromatic degradation are not described in literature, although bacteria belonging to this phylum have been identified in contaminated areas (Feris et al., 2004; Singletonet al., 2011; Sipilä et al., 2008; Silva et al., 2012; Sutton et al., 2013). In agreement, catabolic genes detected on the microarray are related to those observed in Proteobacteria, which dominated CZE soil. Interestingly, however, there was a shift in the catabolic gene structure over time, suggesting that the initial contamination had resulted in the selection of genes encoding benzene/toluene/isopropylbenzene dioxygenases, however, after long-term contamination another signal related catechol 2,3-dioxygenase of *Cupriavidus* sp. BIS7 (WP019451462) increased (Hong et al., 2012). This strains harbors a catabolic gene cluster composed of toluene monooxygenase and phenol monooxygenase encoding genes (WP019451463-75) associated with the extradiol dioxygenase (WP019451462), indicating this strain to degrade benzene via 2 successive monooxygenations. Thus, even though the taxonomical composition of the soils kept constant, the hybridization pattern suggests an adaptation of the metabolic net. Nevertheless, when compared to BRA and SUI microcosm, the reaction in CZE microcosms is minor because of previous long term remediation, as was highlighted before.

In contrast, the bacterial community of the BRA soil reacted dramatically to incubation with benzene, and a clear increase in the relative abundance of Actinobacteria was observed, which correlates with a decrease in the relative abundance of β-Proteobacteria. A similar trend was observed during BTEX incubation, where actinobacterial phylotypes were also enriched. The enrichment of Actinobacteria during incubation of soil with benzene or BTEX was also reflected in the catabolic gene content. In agreement, in the respective microcosms, probes indicating the presence of genes encoding enzymes found in *Rhodococcus* spp strains were observed.

The overall changes at the phylum/class level in SUI soil over time are as rapid as the reactions observed in the BRA microcosms. However, in contrast to BRA benzene treated microcosms, there was an extensive increase in the relative

abundance of γ-proteobacteria during the first 10 days of the experiment. Afterwards, similarly to what was observed in the BRA microcosms, actinobacterial phylotypes increased, together with β-proteobacteria phylotypes. Corroborating with the community data, after 10 days catabolic genes indicating the presence of Pseudomonas strains became abundant, as indicated by signals for the presence of genes encoding naphthalene dioxygenases of *Pseudomonas* strains (AAD02136 and Q51494). In the same way, the enrichment of actinobacterial phylotypes during long-term contamination with benzene, is reflected in the enrichment of actinobacterial catabolic genes such as those related to actinobacterial naphthalene 1,2-dioxygenase (AAD28100) or 2,3-dihydroxybiphenyl dioxygenase (BAA06872).

Overall changes at the phylum/class level in SUI and BRA soil resulted in comparable responses over time, both reactions were rapid. A community succession was observed and after extended incubation, both microcosms undergo an enrichment of actinobacteria. The community succession observed in BRA and SUI microcosms permitted us to conclude that the community reaction when treated with BTEX is slower, compared to the benzene treatment, and enriched phylotypes differed in both microcosms.

In this study, I sequenced for the first time the genome of two representative *Pseudomonas veronii* isolates, *P. veronii* 1YdBTEX2 and 1YB2. Also, I clarified the unique benzene/toluene degradation pathway from *P. veronii*, key players at highly benzene contaminated site. Moreover I could identify some of the genetic attributes that facilitate the growth and persistence of those isolates under harsh conditions: The presence of genes coding for enzymes for the production of the secondary siderophore pyochelin, the presence of a complete denitrification pathway (NO_3^- to N_2) coupled with the presence of a (NiFe) hydrogenase. Those special traits may have enabled these isolates to effective degrade the contaminant and resist in a tricky environment. Additionally, microcosms experiments were set and by the use of high throughput methods, I could characterize the structure of microbial communities and identify catabolic genes, which might be involved in the catabolism of benzene and BTEX and selected under respective pollutant stress. Finally I could link shifts in the catabolic gene structure with shifts in the microbial communities.

5.2 References

Aoki H, Kimura T, Habe H, Yamane H, Kodama T & Omori T (1996) Cloning, nucleotide sequence, and characterization of the genes encoding enzymes involved in the degradation of cumene to 2-hydroxy-6-oxo-7-methylocta-2,4-dienoic acid in Pseudomonas fluorescens IP01. *J Ferment Bioeng* **81***:* 187–196.

Bartels I, Knackmuss HJ & Reineke W (1984) Suicide inactivation of catechol 2,3-dioxygenase from *Pseudomonas putida* mt-2 by 3-halocatechols. *Appl Environ Microbiol* **47**: 500-505.

Berthrong S, Schadt C W, Piñeiro G & Jackson RB (2009) Afforestation alters the composition of functional genes in soil and biogeochemical processes in South American grasslands. *Appl Environ Microbiol* **75**: 6240–6248.

Beil S, Happe B, Timmis KN & Pieper DH (1997) Genetic and biochemical characterization of the broad spectrum chlorobenzene dioxygenase from *Burkholderia* sp. strain PS12 decholorination of 1,2,4,5-tetrachlorobenzene. *Eur J Biochem* **247**: 190-199.

Beil S, Timmis KN & Pieper DH (1999) Genetic and biochemical analyses of the *tec* operon suggest a route for evolution of chlorobenzene degradation genes. *J Bacteriol* **181**: 341-346.

Bohorquez LC, Delgado-Serrano L, López G, Osorio-Forero C, Klepac-Ceraj V, Kolter R, Junca H, Baena S & Zambrano MM (2012) In-depth characterization via complementing culture-independent approaches of the microbial community in an acidic hot spring of the Colombian Andes. *Microbial Ecol* **63**: 103–115.

Camarinha-Silva A, Wos-Oxley ML, Jáuregui R, Becker K & Pieper DH (2012) Validating T-RFLP as a sensitive and high-throughput approach to assess bacterial diversity patterns in human anterior nares. *FEMS Microbiol Ecol* **79**: 98–108.

Corseuil HX, Dal M & Marins M (1997) Contaminação de águas subterrâneas por derramamentos de gasolina: o problema é grave? *Revista Engenharia Sanitária e Ambiental* **2**: 50–54.

Eaton RW (1996) p-Cumate catabolic pathway in *Pseudomonas putida* Fl: cloning and characterization of DNA carrying the *cmt* operon. *J Bacteriol* **178**: 1351–1362.

Eltis LD & Bolin JT (1996) Evolutionary relationships among extradiol dioxygenases. *J Bacteriol* **178**: 5930–5937.

Ettwig KF, Butler MK, Le Paslier D, Pelletier E, Mangenot S, Kuypers MM, Schreiber F, Dutilh BE, Zedelius J, de Beer D, Gloerich J, Wessels HJ, van Alen T,Luesken F, Wu ML, van de Pas-Schoonen KT, Op den Camp HJ, Janssen-Megens EM, Francoijs KJ, Stunnenberg H, Weissenbach J, Jetten MS & Strous M (2010) Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543–548.

Feris KP, Hristova K, Gebreyesus B, Mackay D & Scow KM (2004) A shallow BTEX and MTBE contaminated aquifer supports a diverse microbial community. *Microbial Ecol* **48**: 589–600.

Fong KP, Goh CB & Tan HM (1996) Characterization and expression of the plasmid-borne *bed*D gene from *Pseudomonas putida* ML2, which codes for NAD+-dependent cis-benzene dihydrodiol dehydrogenase. *J Bacteriol* **178**: 5592-5601.

Harayama S, Mermod N, Rekik M, Lehrbach PR & Timmis KN (1987) Roles of divergent branches of the *meta*-cleavage pathway in the degradation of benzoate and substituted benzoates. *J Bacteriol* **169**: 558-564.

Haroon MF, Hu S, Shi Y, Imelfort M, Keller J, Hugenholtz P, Yuan Z & Tyson GW (2013) Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**: 567–570.

He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL, Li X, Wu L, Gentry TJ, Yin Y, Liebich J, Hazen TC & Zhou J (2010) GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J* **4**: 1167–1179.

Hong KW, Thinagaran DAl, Gan HM, Yin W-F & Chan KG (2012) Whole-genome sequence of *Cupriavidus* sp. strain BIS7, a heavy-metal-resistant bacterium. *J Bacteriol* **194**: 6324.

Junca H & Pieper DH (2003) Amplified functional DNA restriction analysis to determine catechol 2,3-dioxygenase gene diversity in soil bacteria. *J Microbiol Methods* **55**: 697–708.

Junca H & Pieper DH (2004) Functional gene diversity analysis in BTEX contaminated soils by means of PCR-SSCP DNA fingerprinting: comparative diversity assessment against bacterial isolates and PCR-DNA clone libraries. *Environ Microbiol* **6**: 95–110.

Junca H, Plumeier I, Hecht H-J & Pieper DH (2004) Difference in kinetic behaviour of catechol 2,3-dioxygenase variants from a polluted environment. *Microbiology* **150**: 4181–4187.

Lapertot M, Seignez C, Ebrahimi S, Delorme S & Peringer P (2007) Mass production of bacterial communities adapted to the degradation of volatile organic compounds (TEX). *Biodegradation* **18**: 343–350.

Loper JE, Hassan KA, Mavrodi DV, Davis EW 2nd, Lim CK, Shaffer BT, Elbourne LD, Stockwell VO, Hartney SL, Breakwell K, Henkels MD, Tetu SG, Rangel LI,Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS 3rd, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE & Paulsen IT (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* **8**: e1002784.

Oosterkamp MJ, Veuskens T, Talarico Saia F, Weelink SA, Goodwin LA, Daligault HE, Bruce DC, Detter JC, Tapia R, Han CS, Land ML, Hauser LJ,Langenhoff AA, Gerritse J, van Berkel WJ, Pieper DH, Junca H, Smidt H, Schraa G, Davids M, Schaap PJ, Plugge CM & Stams AJ. Genome analysis and physiological comparison of *Alicycliphilus denitrificans* strains BC and K601(T.) (2013) *PLoS One* **8**: e66971.

Perez-Pantoja, D, Donoso R, Junca H, Gonzalez B & Pieper DH (2009) Phylogenomics of aerobic bacterial degradation of aromatics. In *Handbook of Hydrocarbon and Lipid Microbiology*. Timmis, KN (ed.). Berlin Heidelberg, Springer-Verlag, 1356-1397.

Ramette A, Frapolli M, Saux MF-L, Gruffaz C, Meyer J-M, Défago G, Sutra L & Moënne-Loccoz Y (2011) *Pseudomonas protegens* sp. nov., widespread plant-protecting bacteria producing the biocontrol compounds 2,4-diacetylphloroglucinol and pyoluteorin. *Syst Appl Microbiol* **34**: 180–188.

Redondo-Nieto M, Barret M, Morrisey JP, Germaine K, Martínez-Granero F, Barahona E, Navazo A, Sánchez-Contreras M, Moynihan JA, Giddens SR,Coppoolse ER, Muriel C, Stiekema WJ, Rainey PB, Dowling D, O'Gara F, Martín M & Rivilla R (2012) Genome sequence of the biocontrol strain *Pseudomonas fluorescens* F113. *J Bacteriol* **194**: 1273–1274.

Shim H, Shin E & Yang S-T (2002) A continuous fibrous-bed bioreactor for BTEX biodegradation by a co-culture of *Pseudomonas putida* and *Pseudomonas fluorescens*. *Adv Environ Res* **7**: 203–216.

Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, Knight CG, Malone JG,Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS, Bentley SD, Hothersall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB & Thomson NR (2009) Genomic and genetic analyses of diversity and plant interactions of P*seudomonas fluorescens*. *Genome Biol* **10**: R51.

Silby MW, Winstanley C, Godfrey SAC, Levy SB & Jackson RW (2011) Pseudomonas genomes: diverse and adaptable. *FEMS Microbiol Rev* **35**: 652–680.

Silva CC, Hayden H, Sawbridge T, Mele P, Kruger RH, Rodrigues MV, Costa GG, Vidal RO, Sousa MP, Torres AP, Santiago VM & Oliveira VM (2012) Phylogenetic and functional diversity of metagenomic libraries of phenol degrading sludge from petroleum refinery wastewater treatment system. *AMB Express* **2**: 18.

Singleton DR, Richardson SD & Aitken MD (2011) Pyrosequence analysis of bacterial communities in aerobic bioreactors treating polycyclic aromatic hydrocarbon-contaminated soil. *Biodegradation* **22**: 1061–1073.

Sipilä TP, Keskinen A-K, Akerman M-L, Fortelius C, Haahtela K & Yrjälä K (2008) High aromatic ring-cleavage diversity in birch rhizosphere: PAH treatment-specific changes of I.E.3 group extradiol dioxygenases and 16S rRNA bacterial communities in soil. *ISME J* **2**: 968–981.

Sutton NB, Maphosa F, Morillo JA, Abu Al-Soud W, Langenhoff AA, Grotenhuis T, Rijnaarts HH & Smidt H (2013) Impact of long-term diesel contamination on soil microbial community structure. *Appl Environ Microbiol* **79**: 619–630.

Tan HM & Mason JR (1990) Cloning and expression of the plasmid-encoded benzene dioxygenase genes from *Pseudomonas putida* ML2. *FEMS Microbiol Lett* **60**: 259-264.

Tan HM & Fong KP (1993) Molecular analysis of the plasmid-borne *bed* gene cluster from *Pseudomonas putida* ML2 and cloning of the *cis*-benzene dihydrodiol dehydrogenase gene. *Can J Microbiol* **39**: 357-362.

Tribelli PM, Raiger Iustman LJ, Catone MV, Di Martino C, Revale S, Méndez BS & López NI (2012) Genome sequence of the polyhydroxybutyrate producer *Pseudomonas extremaustralis*, a highly stress-resistant Antarctic bacterium. *J Bacteriol* **194**: 2381–2382.

Vilchez-Vargas R, Geffers R, Suárez-Diez M, Conte I, Waliczek A, Kaser VS, Kralova M, Junca H & Pieper DH (2013) Analysis of the microbial gene landscape and transcriptome for aromatic pollutants and alkane degradation using a novel internally calibrated microarray system. *Environ Microbiol* **15**: 1016–1039.

Weelink SA, Tan NC, ten Broeke H, van den Kieboom C, van Doesburg W, Langenhoff AA, Gerritse J, Junca H & Stams AJ (2008) Isolation and characterization of *Alicycliphilus denitrificans* strain BC, which grows on benzene with chlorate as the electron acceptor. *Appl Environ Microbiol* **74**: 6672-6681.

Witzig R, Junca H, Hecht HJ & Pieper DH (2006) Assessment of toluene/biphenyl dioxygenase gene diversity in benzene-polluted soils: Links between benzene biodegradation and genes similar to those encoding isopropylbenzene dioxygenases. *Appl Environ Microbiol* **72**: 3504–3514.

6. Appendix

Table S1: Phylogenetic assignment of all 501 phylotypes determined using Illumina sequencing.

Phylotype	Length	Designation	Phylogenetic affiliation based on 16S rRNA gene sequence				
			Phylum	Class	Order	Family	Genus
Phy1	86	Hydrogenophilaceae	Proteobacteria	Betaproteobacteria	Hydrogenophilales	Hydrogenophilaceae	Thiobacillus
Phy2	86	Burkholderiaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Burkholderia
Phy3	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	ND
Phy4	86	Micrococcineae	Actinobacteria	Actinobacteria	Actinomycetales	ND	ND
Phy5	86	Rhodocyclaceae	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	Denitratisoma
Phy6	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy7	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy8	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy9	86	Sinobacteraceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae	Hydrocarboniphaga
Phy10	86	Cyclobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cyclobacteriaceae	Algoriphagus
Phy11	86	Comamonadaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	ND
Phy12	86	Burkholderiaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Ralstonia/Burkholderia
Phy13	86	Nocardiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Nocardia
Phy14	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Pseudoxanthomonas
Phy15	86	Mycobacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium
Phy16	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy17	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy18	86	Micrococcaceae	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	ND
Phy19	86	Burkholderiales incertae sedis	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiales incertae sedis	ND
Phy20	86	Chromatiales	Proteobacteria	Gammaproteobacteria	Chromatiales	ND	ND
Phy21	86	Burkholderiaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Limnobacter
Phy22	86	Rhodobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhodobiaceae	Parvibaculum
Phy23	86	Alcaligenaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	ND
Phy24	86	Moraxellaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Perlucidibaca
Phy25	86	Comamonadaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	ND

Phy26	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy27	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy28	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy29	86	Holophagaceae	Acidobacteria	Holophagae	Holophagales	Holophagaceae	Geothrix
Phy30	86	Nocardiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Rhodococcus
Phy31	86	Halothiobacillaceae	Proteobacteria	Gammaproteobacteria	Chromatiales	Halothiobacillaceae	ND
Phy32	86	Comamonadaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Hydrogenophaga
Phy33	86	Cytophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cytophagaceae	Cytophaga
Phy34	86	Comamonadaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Macromonas
Phy35	86	Cellulomonadaceae	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	Cellulomonas
Phy36	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy37	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy38	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy39	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	ND
Phy40	86	Cyclobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cyclobacteriaceae	Algoriphagus
Phy41	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	ND	ND
Phy42	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy43	86	Methylophilaceae	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Methylobacillus
Phy44	86	Nocardiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	ND
Phy45	86	Methylophilaceae	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	ND
Phy46	86	Sinobacteraceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae	ND
Phy47	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy48	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy49	86	Oxalobacteraceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Herminiimonas
Phy50	86	Nocardioidaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae	Nocardioides
Phy51	86	Burkholderiales incertae sedis	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiales incertae sedis	Thiomonas
Phy52	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy53	86	Sphingobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae	ND
Phy54	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium
Phy55	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Lacibacter
Phy56	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium
Phy57	86	Nocardioidaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae	Nocardioides
Phy58	86	Chloroflexi	Chloroflexi	ND	ND	ND	ND
Phy59	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	unclassified Xanthomonadaceae
Phy60	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Arenimonas
Phy61	86	Hydrogenophilaceae	Proteobacteria	Betaproteobacteria	Hydrogenophilales	Hydrogenophilaceae	Thiobacillus
Phy62	86	Bacteroidetes	Bacteroidetes	unclassified Bacteroidetes	ND	ND	ND
Phy63	86	Methylophilaceae	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	ND
Phy64	86	Burkholderiaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Cupriavidus

Phy65	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy66	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Flavobacterium
Phy67	86	Bacteria	ND	ND	ND	ND	ND
Phy68	86	Sphingobacteriales	Bacteroidetes	Sphingobacteria	Sphingobacteriales	ND	ND
Phy69	86	Sphingobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
Phy70	86	Bacillales	Firmicutes	Bacilli	Bacillales	ND	ND
Phy71	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy72	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy73	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Chryseobacterium
Phy74	86	Nocardiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Rhodococcus
Phy75	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy76	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy77	86	Bacteria	ND	ND	ND	ND	ND
Phy78	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy79	86	Sphingobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae	Pedobacter
Phy80	86	Bacteria	ND	ND	ND	ND	ND
Phy81	86	Iamiaceae	Actinobacteria	Actinobacteria	Acidimicrobiales	Iamiaceae	Iamia
Phy82	86	Conexibacteraceae	Actinobacteria	Actinobacteria	ND	Conexibacteraceae	Conexibacter
Phy83	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Arenimonas
Phy84	86	Intrasporangiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	ND
Phy85	86	Corynebacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium
Phy86	86	Rhodospirillaceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	Oceanibaculum
Phy87	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy88	86	Bacilli	Firmicutes	Bacilli	Bacillales	ND	ND
Phy89	86	Micrococcinae	Actinobacteria	Actinobacteria	Actinomycetales	ND	ND
Phy90	86	Nocardioidaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae	ND
Phy91	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy92	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy93	86	Bacteroidetes	Bacteroidetes	unclassified Bacteroidetes	ND	ND	ND
Phy94	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
Phy95	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Novosphingobium
Phy96	86	Porphyromonadaceae	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	ND
Phy97	86	Bacillaceae	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
Phy98	86	Gemmatimonadaceae	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
Phy99	86	Methylobacteriaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium
Phy100	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy101	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	ND
Phy102	86	Comamonadaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Polaromonas
Phy103	86	Methanosarcinales	Euryarchaeota	Methanomicrobia	Methanosarcinales	ND	ND

Phy104	86	Sinobacteraceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae	Nevskia
Phy105	86	Caulobacteraceae	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	ND
Phy106	86	Chloroflexi	Chloroflexi	ND	ND	ND	ND
Phy107	86	Caulobacteraceae	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Brevundimonas
Phy108	86	Gp1	Acidobacteria	Acidobacteria Gp1	ND	ND	Gp1
Phy109	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	ND	ND
Phy110	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Ferruginibacter
Phy111	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Chitinophaga
Phy112	86	Actinobacteria	Actinobacteria	Actinobacteria	ND	ND	ND
Phy113	86	Nocardioidaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae	Nocardioides
Phy114	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy115	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Pseudoxanthomonas
Phy116	86	Bradyrhizobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	ND
Phy117	86	Burkholderiales	Proteobacteria	Betaproteobacteria	Burkholderiales	ND	ND
Phy118	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	ND
Phy119	86	Erythrobacteraceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Erythrobacteraceae	Erythrobacter
Phy120	86	Comamonadaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Ramlibacter
Phy121	86	Chloroflexi	Chloroflexi	ND	ND	ND	ND
Phy122	86	Sphingomonadales	Proteobacteria	Alphaproteobacteria	Sphingomonadales	ND	ND
Phy123	86	Actinomycetales	Actinobacteria	Actinobacteria	Actinomycetales	ND	ND
Phy124	86	Thermoprotei	Crenarchaeota	Thermoprotei	ND	ND	ND
Phy125	86	Porphyromonadaceae	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides
Phy126	86	Bacteriovoracaceae	Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	Peredibacter
Phy127	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium
Phy128	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy129	86	Thermoprotei	Crenarchaeota	Thermoprotei	ND	ND	ND
Phy130	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy131	86	Rhodospirillaceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	ND
Phy132	86	Hydrogenophilaceae	Proteobacteria	Betaproteobacteria	Hydrogenophilales	Hydrogenophilaceae	Thiobacillus
Phy133	86	OD1	OD1	ND	ND	ND	ND
Phy134	86	Desulfobulbaceae	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfobulbaceae	ND
Phy135	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy136	86	Iamiaceae	Actinobacteria	Actinobacteria	Acidimicrobiales	Iamiaceae	Iamia
Phy137	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy138	86	Methylophilaceae	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	ND
Phy139	86	Acidimicrobidae	Actinobacteria	Actinobacteria	ND	ND	ND
Phy140	86	Anaerolineaceae	Chloroflexi	Anaerolineae	Anaerolineales	Anaerolineaceae	ND
Phy141	86	Caulobacteraceae	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Asticcacaulis
Phy142	86	Gemmatimonadetes	Gemmatimonadetes	ND	ND	ND	ND

Phy143	86	Bacteroidetes	Bacteroidetes	ND	ND	ND	ND
Phy144	86	Peptococcaceae	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	ND
Phy145	86	Bacteria	ND	ND	ND	ND	ND
Phy146	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy147	86	Bacteria	ND	ND	ND	ND	ND
Phy148	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Rathayibacter
Phy149	86	Sporichthyaceae	Actinobacteria	Actinobacteria	Actinomycetales	Sporichthyaceae	Sporichthya
Phy150	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy151	86	Comamonadaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	ND
Phy152	86	Solirubrobacterales	Actinobacteria	Actinobacteria	Solirubrobacterales	ND	ND
Phy153	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	ND
Phy154	86	Actinobacteria	Actinobacteria	Actinobacteria	ND	ND	ND
Phy155	86	Bacteria	ND	ND	ND	ND	ND
Phy156	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Flavobacterium
Phy157	86	Propionibacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	ND
Phy158	86	Spirochaetaceae	Spirochaetes	Spirochaetes	Spirochaetales	Spirochaetaceae	Treponema
Phy159	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	ND
Phy160	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Arenimonas
Phy161	86	Micrococcaceae	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Arthrobacter
Phy162	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Arenimonas
Phy163	86	Mycobacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium
Phy164	86	Bacteria	ND	ND	ND	ND	ND
Phy165	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy166	86	Oxalobacteraceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	ND
Phy167	86	Clostridiales	Firmicutes	Clostridia	Clostridiales	ND	ND
Phy168	86	Parachlamydiaceae	Chlamydiae	Chlamydiae	Chlamydiales	Parachlamydiaceae	ND
Phy169	86	Bacteria	ND	ND	ND	ND	ND
Phy170	86	Acidimicrobidae incertae sedis	Actinobacteria	Actinobacteria	Acidimicrobidae incertae sedis	ND	Ilumatobacter
Phy171	86	Bacteroidetes incertae sedis	Bacteroidetes	ND	ND	Bacteroidetes incertae sedis	Prolixibacter
Phy172	86	Legionellaceae	Proteobacteria	Gammaproteobacteria	Legionellales	Legionellaceae	Legionella
Phy173	86	Methanosaetaceae	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosaetaceae	Methanosaeta
Phy174	86	Solirubrobacterale	Actinobacteria	Actinobacteria	Solirubrobacterale	ND	ND
Phy175	86	Thermomicrobia	Chloroflexi	Thermomicrobia	ND	ND	ND
Phy176	86	Spirochaetaceae	Spirochaetes	Spirochaetes	Spirochaetales	Spirochaetaceae	ND
Phy177	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy178	86	Geobacteraceae	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	Geobacteraceae	Geobacter
Phy179	86	Solirubrobacteraceae	Actinobacteria	Actinobacteria	Solirubrobacterales	Solirubrobacteraceae	Solirubrobacter
Phy180	86	Sphingobacteriales	Bacteroidetes	Sphingobacteria	Sphingobacteriales	ND	ND
Phy181	86	Bacteria	ND	ND	ND	ND	ND

Phy182	86	Ruminococcaceae	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Acetivibrio
Phy183	86	Conexibacteraceae	Actinobacteria	Actinobacteria	Solirubrobacterales	Conexibacteraceae	Conexibacter
Phy184	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy185	86	Clostridia	Firmicutes	Clostridia	ND	ND	ND
Phy186	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy187	86	Bacteria	ND	ND	ND	ND	ND
Phy188	86	Bacteria	ND	ND	ND	ND	ND
Phy189	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy190	86	Acidimicrobidae	Actinobacteria	Actinobacteria	ND	ND	ND
Phy191	86	Bacteria	ND	ND	ND	ND	ND
Phy192	86	Sphingobacteriales	Bacteroidetes	Sphingobacteria	Sphingobacteriales	ND	ND
Phy193	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy194	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	ND
Phy195	86	Bacteroidetes	Bacteroidetes	ND	ND	ND	ND
Phy196	86	Solirubrobacterales	Actinobacteria	Actinobacteria	Solirubrobacterales	ND	ND
Phy197	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy198	86	OD1	OD1	ND	ND	ND	OD1 genera incertae sedis
Phy199	86	Burkholderiaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	ND
Phy200	86	Burkholderiales	Proteobacteria	Betaproteobacteria	Burkholderiales	ND	ND
Phy201	86	Sphingobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae	Mucilaginibacter
Phy202	86	Bacteroidetes	Bacteroidetes	ND	ND	ND	ND
Phy203	86	Bacteria	ND	ND	ND	ND	ND
Phy204	86	Sphingomonadales	Proteobacteria	Alphaproteobacteria	Sphingomonadales	ND	ND
Phy205	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy206	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy207	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy208	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Yeosuana
Phy209	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy210	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	ND
Phy211	86	Legionellaceae	Proteobacteria	Gammaproteobacteria	Legionellales	Legionellaceae	Legionella
Phy212	86	Oxalobacteraceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	ND
Phy213	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy214	86	Desulfuromonadales	Proteobacteria	Deltaproteobacteria	Desulfuromonadales	ND	ND
Phy215	86	Thermoactinomycetaceae	Firmicutes	Bacilli	Bacillales	Thermoactinomycetaceae	Planifilum
Phy216	86	Bacteria	ND	ND	ND	ND	ND
Phy217	86	Solirubrobacterales	Actinobacteria	Actinobacteria	Solirubrobacterales	ND	ND
Phy218	86	Enterobacteriaceae	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	ND
Phy219	86	Moraxellaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Perlucidibaca
Phy220	86	Bacteria	ND	ND	ND	ND	ND

Phy221	86	Cellulomonadaceae	Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	ND
Phy222	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy223	86	OD1	OD1	ND	ND	ND	OD1 genera incertae sedis
Phy224	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	ND
Phy225	86	Sphingobacteriales	Bacteroidetes	Sphingobacteria	Sphingobacteriales	ND	ND
Phy226	86	Hyphomicrobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Pedomicrobium
Phy227	86	Burkholderiales incertae sedis	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiales incertae sedis	ND
Phy228	86	Conexibacteraceae	Actinobacteria	Actinobacteria	Solirubrobacterales	Conexibacteraceae	ND
Phy229	86	OD1	OD1	ND	ND	ND	OD1 genera incertae sedis
Phy230	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	ND
Phy231	86	Nocardioidaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae	Nocardioides
Phy232	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy233	86	Sinobacteraceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae	Singularimonas
Phy234	86	Gammaproteobacteriaincertaesedis	Proteobacteria	Gammaproteobacteria	ND	Gammaproteobacteria incertae sedis	Methylohalomonas
Phy235	86	Bacillaceae	Firmicutes	Bacilli	Bacillales	Bacillaceae	ND
Phy236	86	Sphaerobacteraceae	Chloroflexi	Thermomicrobia	Sphaerobacterales	Sphaerobacteraceae	Sphaerobacter
Phy237	86	Kineosporiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Kineosporiaceae	ND
Phy238	86	Thermomicrobia	Chloroflexi	Thermomicrobia	ND	ND	ND
Phy239	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
Phy240	86	Sphaerobacteraceae	Chloroflexi	Thermomicrobia	Sphaerobacterales	Sphaerobacteraceae	Sphaerobacter
Phy241	86	Hyphomicrobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	ND
Phy242	86	Actinomycetales	Actinobacteria	Actinobacteria	Actinomycetales	ND	ND
Phy243	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Terrimonas
Phy244	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy245	86	Bacteria	ND	ND	ND	ND	ND
Phy246	86	OD1	OD1	ND	ND	ND	OD1 genera incertae sedis
Phy247	86	Legionellaceae	Proteobacteria	Gammaproteobacteria	Legionellales	Legionellaceae	Legionella
Phy248	86	Rhodocyclaceae	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	ND
Phy249	86	Pseudonocardiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	Pseudonocardia
Phy250	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy251	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Cohnella
Phy252	86	Gemmatimonadaceae	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
Phy253	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy254	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	ND
Phy255	86	Bacteroidetes	Bacteroidetes	ND	ND	ND	ND
Phy256	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy257	86	Cystobacteraceae	Proteobacteria	Deltaproteobacteria	Myxococcales	Cystobacteraceae	Anaeromyxobacter
Phy258	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Roseomonas
Phy259	86	Bacteria	ND	ND	ND	ND	ND

Phy260	86	Thermoactinomycetaceae	Firmicutes	Bacilli	Bacillales	Thermoactinomycetaceae	Planifilum
Phy261	86	Bacillaceae	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
Phy262	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy263	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy264	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	ND
Phy265	86	Sphingobacteriales	Bacteroidetes	Sphingobacteria	Sphingobacteriales	ND	ND
Phy266	86	Verrucomicrobiaceae	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	Prosthecobacter
Phy267	86	Sphingobacteriales	Bacteroidetes	Sphingobacteria	Sphingobacteriales	ND	ND
Phy268	86	Chromatiaceae	Proteobacteria	Gammaproteobacteria	Chromatiales	Chromatiaceae	Rheinheimera
Phy269	86	Deltaproteobacteria	Proteobacteria	Deltaproteobacteria	ND	ND	ND
Phy270	86	Bacteria	ND	ND	ND	ND	ND
Phy271	86	Hyphomicrobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Devosia
Phy272	86	Erysipelotrichaceae	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	ND
Phy273	86	Moraxellaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter
Phy274	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy275	86	Clostridiales	Firmicutes	Clostridia	Clostridiales	ND	ND
Phy276	86	Planctomycetaceae	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae	Planctomyces
Phy277	86	Chlamydiae	Chlamydiae	ND	ND	ND	ND
Phy278	86	Chloroflexi	Chloroflexi	ND	ND	ND	ND
Phy279	86	Bacillaceae	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
Phy280	86	Acidimicrobidae	Actinobacteria	Actinobacteria	ND	ND	ND
Phy281	86	Bacteria	ND	ND	ND	ND	ND
Phy282	86	Bacteroidetes	Bacteroidetes	ND	ND	ND	ND
Phy283	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy284	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Dokdonella
Phy285	86	Planctomycetes	Planctomycetes	ND	ND	ND	ND
Phy286	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy287	86	Cyclobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cyclobacteriaceae	ND
Phy288	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy289	86	Burkholderiaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	ND
Phy290	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy291	86	Sphingobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae	Mucilaginibacter
Phy292	86	Veillonellaceae	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	ND
Phy293	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Cellvibrio
Phy294	86	Nitrospiraceae	Nitrospira	Nitrospira	Nitrospirales	Nitrospiraceae	ND
Phy295	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	ND
Phy296	86	Nocardioidaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae	ND
Phy297	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy298	86	Bacteria	ND	ND	ND	ND	ND

Phy299	86	Thermoprotei	Crenarchaeota	Thermoprotei	ND	ND	ND
Phy300	86	Bacteria	ND	ND	ND	ND	ND
Phy301	86	Acidimicobiales	Actinobacteria	Actinobacteria	Acidimicobiales	ND	ND
Phy302	86	Sphingobacteriales	Bacteroidetes	Sphingobacteria	Sphingobacteriales	ND	ND
Phy303	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Azotobacter
Phy304	86	Clostridia	Firmicutes	Clostridia	ND	ND	ND
Phy305	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy306	86	Rhodospirillaceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	ND
Phy307	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy308	86	Bacteroidetes	Bacteroidetes	ND	ND	ND	ND
Phy309	86	Micromonosporaceae	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	ND
Phy310	86	Parachlamydiaceae	Chlamydiae	Chlamydiae	Chlamydiales	Parachlamydiaceae	ND
Phy311	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Stella
Phy312	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	ND
Phy313	86	IncertaeSedisXVIII	Firmicutes	Clostridia	Clostridiales	Incertae Sedis XVIII	Symbiobacterium
Phy314	86	Erysipelotrichaceae	Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	ND
Phy315	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy316	86	Hyphomicrobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	ND
Phy317	86	Bacillaceae	Firmicutes	Bacilli	Bacillales	Bacillaceae	Exiguobacterium
Phy318	86	Methanosarcinaceae	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae	Methanosarcina
Phy319	86	Gracilibacteraceae	Firmicutes	Clostridia	Clostridiales	Gracilibacteraceae	ND
Phy320	86	Bradyrhizobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Afipia
Phy321	86	Hyphomicrobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Devosia
Phy322	86	Chloroflexi	Chloroflexi	ND	ND	ND	ND
Phy323	86	Sorangiineae	Proteobacteria	Deltaproteobacteria	Myxococcales	Sorangiineae	ND
Phy324	86	Sphingobacteriales	Bacteroidetes	Sphingobacteria	Sphingobacteriales	ND	ND
Phy325	86	Rhizobiales	Proteobacteria	Alphaproteobacteria	Rhizobiales	ND	ND
Phy326	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy327	86	Halomonadaceae	Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	Halomonas
Phy328	86	GP3	Acidobacteria	Acidobacteria Gp3	ND	ND	ND
Phy329	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy330	86	Clostridiales	Firmicutes	Clostridia	Clostridiales	ND	ND
Phy331	86	Burkholderiaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Burkholderia
Phy332	86	Alcaligenaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	ND
Phy333	86	Cyclobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cyclobacteriaceae	Algoriphagus
Phy334	86	Nitrosomonadaceae	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae	Nitrosospira
Phy335	86	Trueperaceae	Deinococcus-Thermus	Deinococci	Deinococcales	Trueperaceae	Truepera
Phy336	86	Firmicutes	Firmicutes	ND	ND	ND	ND
Phy337	86	Verrucomicrobiaceae	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	ND

Phy338	86	Acidimicrobineae	Actinobacteria	Actinobacteria	Acidimicrobiales	ND	ND
Phy339	86	Planctomycetaceae	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae	Rhodospirellula
Phy340	86	Bacteria	ND	ND	ND	ND	ND
Phy341	86	Sphingobacteriaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriaceae	ND
Phy342	86	Acidimicrobineae	Actinobacteria	Actinobacteria	Acidimicrobiales	ND	ND
Phy343	86	Desulfobulbaceae	Proteobacteria	Deltaproteobacteria	Desulfobacterales	Desulfobulbaceae	ND
Phy344	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriaceae	ND	ND
Phy345	86	Solirubrobacterales	Actinobacteria	Actinobacteria	Solirubrobacterales	ND	ND
Phy346	86	Bacteria	ND	ND	ND	ND	ND
Phy347	86	Micrococcineae	Actinobacteria	Actinobacteria	Actinomycetales	ND	ND
Phy348	86	Sphaerobacteraceae	Chloroflexi	Thermomicrobia	Sphaerobacterales	Sphaerobacteraceae	Sphaerobacter
Phy349	86	Planctomycetaceae	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae	Pirellula
Phy350	86	Caulobacteraceae	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Brevundimonas
Phy351	86	Rhodocyclaceae	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	Methyloversatilis
Phy352	86	Thermoactinomycetaceae	Firmicutes	Bacilli	Bacillales	Thermoactinomycetaceae	ND
Phy353	86	Actinobacteria	Actinobacteria	ND	ND	ND	ND
Phy354	86	Rhizobiales	Proteobacteria	Alphaproteobacteria	Rhizobiales	ND	ND
Phy355	86	IncertaeSedisXVIII	Firmicutes	Clostridia	Clostridiales	Incertae Sedis XVIII	Symbiobacterium
Phy356	86	Rhodospirillaceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	ND
Phy357	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
Phy358	86	Gemmatimonadaceae	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
Phy359	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	ND
Phy360	86	Trueperaceae	Deinococcus-Thermus	Deinococci	Deinococcales	Trueperaceae	Truepera
Phy361	86	Actinobacteria	Actinobacteria	Actinobacteria	ND	ND	ND
Phy362	86	Moraxellaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Perlucidibaca
Phy363	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Brevibacillus
Phy364	86	Actinomycetales	Actinobacteria	Actinobacteria	Actinomycetales	ND	ND
Phy365	86	Rhodospirillaceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	ND
Phy366	86	Rhodospirillales	Proteobacteria	Alphaproteobacteria	Rhodospirillales	ND	ND
Phy367	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	ND
Phy368	86	Rhizobiales	Proteobacteria	Alphaproteobacteria	Rhizobiales	ND	ND
Phy369	86	Methylophilaceae	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	ND
Phy370	86	Thermoactinomycetaceae	Firmicutes	Bacilli	Bacillales	Thermoactinomycetaceae	Thermoactinomyces
Phy371	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidocella
Phy372	86	Thermoactinomycetaceae	Firmicutes	Bacilli	Bacillales	Thermoactinomycetaceae	Thermoflavimicrobium
Phy373	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Flavobacterium
Phy374	86	Bacilli	Firmicutes	Bacilli	ND	ND	ND
Phy375	86	Bacillales	Firmicutes	Bacilli	Bacillales	ND	ND
Phy376	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Ammoniphilus

Phy377	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy378	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy379	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy380	86	Rhizobiales	Proteobacteria	Alphaproteobacteria	Rhizobiales	ND	ND
Phy381	86	Bacteroidetes incertae sedis	Bacteroidetes	ND	ND	Bacteroidetes incertae sedis	ND
Phy382	86	Acidimicrobinae	Actinobacteria	Actinobacteria	Acidimicobiales	ND	ND
Phy383	86	Patulibacteraceae	Actinobacteria	Actinobacteria	Solirubrobacterales	Patulibacteraceae	Patulibacter
Phy384	86	Rhodospirillaceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	ND
Phy385	86	Nitrosomonadaceae	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae	Nitrosomonas
Phy386	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Curtobacterium
Phy387	86	Burkholderiales incertae sedis	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiales incertae sedis	Thiomonas
Phy388	86	Sphaerobacteraceae	Chloroflexi	Thermomicrobia	Sphaerobacterales	Sphaerobacteraceae	Sphaerobacter
Phy389	86	Intrasporangiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	ND
Phy390	86	Cryomorphaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Cryomorphaceae	Fluviicola
Phy391	86	TM7	TM7	TM7 genera incertae sedis	ND	ND	TM7 genera incertae sedis
Phy392	86	Betaproteobacteria	Proteobacteria	Betaproteobacteria	ND	ND	ND
Phy393	86	Pseudonocardiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	ND
Phy394	86	Actinomycetales	Actinobacteria	Actinobacteria	Actinomycetales	ND	ND
Phy395	86	Bacteria	ND	ND	ND	ND	ND
Phy396	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy397	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy398	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	ND
Phy399	86	Verrucomicrobiaceae	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	Luteolibacter
Phy400	86	Rhodobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	ND
Phy401	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Dokdonella
Phy402	86	Solirubrobacterales	Actinobacteria	Actinobacteria	Solirubrobacterales	ND	ND
Phy403	86	Bacteria	ND	ND	ND	ND	ND
Phy404	86	Cryomorphaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Cryomorphaceae	Owenweeksia
Phy405	86	Verrucomicrobiaceae	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	ND
Phy406	86	Peptococcaceae	Firmicutes	Clostridia	Clostridiales	Peptococcaceae	ND
Phy407	86	Actinobacteria	Actinobacteria	Actinobacteria	ND	ND	ND
Phy408	86	Alcaligenaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Castellaniella
Phy409	86	Rhizobiales	Proteobacteria	Alphaproteobacteria	Rhizobiales	ND	ND
Phy410	86	Bacteria	ND	ND	ND	ND	ND
Phy411	86	Bacillaceae	Firmicutes	Bacilli	Bacillales	Bacillaceae	Anoxybacillus
Phy412	86	Verrucomicrobia	Verrucomicrobia	ND	ND	ND	ND
Phy413	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy414	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy415	86	Conexibacteracea	Actinobacteria	Actinobacteria	Solirubrobacterales	Conexibacteracea	Conexibacter

Phy416	86	Rhodocyclaceae	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	ND
Phy417	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	ND
Phy418	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy419	86	Deltaproteobacteria	Proteobacteria	Deltaproteobacteria	ND	ND	ND
Phy420	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy421	86	Rubrobacteraceae	Actinobacteria	Actinobacteria	Rubrrobacteriales	Rubrobacteraceae	Rubrobacter
Phy422	86	Rhizobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium
Phy423	86	Bacteria	ND	ND	ND	ND	ND
Phy424	86	Veillonellaceae	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	ND
Phy425	86	Rubrobacteraceae	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae	Rubrobacter
Phy426	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	ND
Phy427	86	Bacteria	ND	ND	ND	ND	ND
Phy428	86	Bacteria	ND	ND	ND	ND	ND
Phy429	86	Rhizobiaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium
Phy430	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	ND
Phy431	86	Bacteria	ND	ND	ND	ND	ND
Phy432	86	Parachlamydiaceae	Chlamydiae	Chlamydiae	Chlamydiales	Parachlamydiaceae	Parachlamydia
Phy433	86	Verrucomicrobia	Verrucomicrobia	ND	ND	ND	ND
Phy434	86	Bacteria	ND	ND	ND	ND	ND
Phy435	86	Verrucomicrobia	Verrucomicrobia	ND	ND	ND	ND
Phy436	86	Nocardiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	Gordonia
Phy437	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	ND
Phy438	86	Rhizobiales	Proteobacteria	Alphaproteobacteria	Rhizobiales	ND	ND
Phy439	86	Micromonosporaceae	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	ND
Phy440	86	Spirochaetaceae	Spirochaetes	Spirochaetes	Spirochaetales	Spirochaetaceae	ND
Phy441	86	Kineosporiaceae	Actinobacteria	Actinobacteria	Actinomycetales	Kineosporiaceae	Kineococcus
Phy442	86	Moraxellaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Alkanindiges
Phy443	86	Rhizobiales	Proteobacteria	Alphaproteobacteria	Rhizobiales	ND	ND
Phy444	86	Rhodospirillaceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	ND
Phy445	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Lysobacter
Phy446	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Ferruginibacter
Phy447	86	Bacillaceae	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
Phy448	86	TM7	TM7	ND	ND	ND	TM7 genera incertae sedis
Phy449	86	Hydrogenophilaceae	Proteobacteria	Betaproteobacteria	Hydrogenophilales	Hydrogenophilaceae	Thiobacillus
Phy450	86	Gammaproteobacteria	Proteobacteria	Gammaproteobacteria	ND	ND	ND
Phy451	86	Pseudomonadaceae	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
Phy452	86	Chloroflexi	Chloroflexi	ND	ND	ND	ND
Phy453	86	Bacteroidaceae	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
Phy454	86	Prevotellaceae	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella

Phy455	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	ND
Phy456	86	Microbacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Agrococcus
Phy457	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Lysobacter
Phy458	86	Propionibacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	ND
Phy459	86	Bacteroidetes	Bacteroidetes	ND	ND	ND	ND
Phy460	86	Pseudonocardinae	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardinae	ND
Phy461	86	Sorangiineae	Proteobacteria	Deltaproteobacteria	Myxococcales	Sorangiineae	ND
Phy462	86	Proteobacteria	Proteobacteria	ND	ND	ND	ND
Phy463	86	Comamonadaceae	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Variovorax
Phy464	86	TM7	TM7	TM7 genera incertae sedis	ND	ND	ND
Phy465	86	Legionellaceae	Proteobacteria	Gammaproteobacteria	Legionellales	Legionellaceae	Legionella
Phy466	86	TM7	TM7	TM7 genera incertae sedis	ND	ND	ND
Phy467	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	Terrimonas
Phy468	86	TM7	TM7	TM7 genera incertae sedis	ND	ND	ND
Phy469	86	Rhodocyclaceae	Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	Shinella
Phy470	86	Methylobacteriaceae	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium
Phy471	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy472	86	Chlamydiae	Chlamydiae	Chlamydiae	Chlamydiales	ND	Ilumatobacter
Phy473	86	Acidimicrobidae incertae sedis	Actinobacteria	Actinobacteria	ND	Acidimicrobidae incertae sedis	ND
Phy474	86	Gp4	Acidobacteria	Acidobacteria Gp4	ND	ND	Gp4
Phy475	86	Propionibacteriaceae	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	ND
Phy476	86	Chitinophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Chitinophagaceae	ND
Phy477	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Paenibacillus
Phy478	86	Rhizobiales	Proteobacteria	Alphaproteobacteria	Rhizobiales	ND	ND
Phy479	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy480	86	Prevotellaceae	Bacteroidetes	Bacteroidia	Bacteroidiales	prevotellaceae	ND
Phy481	86	Xanthomonadaceae	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Luteimonas
Phy482	86	Myxococcales	Proteobacteria	Deltaproteobacteria	Myxococcales	ND	ND
Phy483	86	Clostridiaceae	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	ND
Phy484	86	Alphaproteobacteria	Proteobacteria	Alphaproteobacteria	ND	ND	ND
Phy485	86	Paenibacillaceae	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	Thermobacillus
Phy486	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	ND
Phy487	86	Spartobacteria genera incertae sedis	Verrucomicrobia	Spartobacteria	ND	Spartobacteria genera incertae sedis	ND
Phy488	86	Acetobacteraceae	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidisphaera
Phy489	86	Proteobacteria	Proteobacteria	ND	ND	ND	ND
Phy490	86	Verrucomicrobiaceae	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	Haloferula
Phy491	86	Nitrospiraceae	Nitrospira	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira
Phy492	86	Cytophagaceae	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Cytophagaceae	Dyadobacter
Phy493	86	Gp4	Acidobacteria	Acidobacteria Gp4	ND	ND	Gp4

Phy494	86	Sphingomonadaceae	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
Phy495	86	Porphyromonadaceae	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
Phy496	86	Gp4	Acidobacteria	Acidobacteria Gp4	ND	ND	Gp4
Phy497	86	Gp4	Acidobacteria	Acidobacteria Gp4	ND	ND	Gp4
Phy498	86	Prevotellaceae	Bacteroidetes	Bacteroidia	Bacteroidales	prevotellaceae	ND
Phy499	86	Flavobacteriaceae	Bacteroidetes	Flavobacteria	Flavobacteriales	Flavobacteriaceae	Capnocytophaga
Phy500	86	Pasteurellaceae	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	ND
Phy501	86	Fusobacteriaceae	Fusobacteria	Fusobacteria	Fusobacteriales	Fusobacteriaceae	Fusobacterium

*ND: not determined

Curriculum Vitae

Daiana de Lima Morales

E-mail: daia.morales@gmail.com

Date and Place of Birth: August 3rd 1981, Porto Alegre, Brazil

Education

April.2009-May.2014	**Doctoral Student** Helmholtz Centre for Infection Research (HZI) Department of Microbial Interactions and Processes (Prof. Dr. Dietmar H. Pieper)
March.2006-May.2008	**Master Degree** Federal University of Rio Grande do Sul (UFRGS) Department of Agricultural and Environmental (Prof. Dr. Fátima Menezes Bento)
March.2005-July.2007	**Master of Business Administration - MBA** Foundation Getúlio Vargas (FGV) Environmental Management
March.2001-July.2005	**Diplom in Biology** Catholic University from Rio Grande do Sul (PUCRS)
March.2003-July.2005	**Research in Microbiology** Federal University of Rio Grande do Sul (UFRGS) Department of Agricultural and Environmental (Prof. Dr. Gertrudes Corção)

PUBLICATIONS

Einsfeld A, Canal N, **Morales D**, Lorentz R, Fuentefria D B, Corção G (2007) Characterization of the enterocins produced by *Enterococcus mundtii* isolated from humans feces. *Braz Arch Biol Technol* **50**: 249-258**.**

de Lima-Morales D, Chaves-Moreno D, Jarek M, Vilchez-Vargas R, Jauregui R & Pieper DH (2013) Draft genome sequence of *Pseudomonas veronii* strain 1YdBTEX2. *Genome Announc* **1**: e00258-13.

SCHOLARSHIPS

April.2009-Sept.2013	**Doctoral Scholarship** Brazilian National Council for Scientific and Technological Development (CNPq)
March.2006-May.2008	**Master Scholarship** Brazilian National Council for Scientific and Technological Development (CNPq)
March.2004-July.2005	**Scholarship for Research Internship** Foundation for Research Support of the State of Rio Grande do Sul (FAPERGS)

LANGUAGE SKILLS

Fluent in Portuguese, English and basic German and Spanish

CONFERENCE CONTRIBUTIONS

Morales D, Chaves D, Vilchez-Vargas R, Jáuregui R and Pieper DH (2012) The genome of novel benzene degrading *Pseudomonas veronii* strains. (Oral Presentation) Pollutant biodegradation under environmental stress: towards rational bioaugmentation – Amsterdam – The Netherlands

Morales D, Chaves D, Vilchez-Vargas R, Jáuregui R & Pieper DH (2011) The genome of two *Pseudomonas veronii* Strains Crucial for Benzene degradation in a Benzene Contaminated Site. (Poster) ProkaGENOMICS – Göttingen – Deutschland

Morales D, Vilchez-Vargas R & Pieper DH (2010) Catabolic gene landscape of novel benzene degrading Pseudomonas strains. (Oral Presentation) DGHM/VAAM– Hannover – Deutschland

Morales D, Passos C, Weissheimer MIK, Jacques RJS, Peralba MCR & Bento FM (2007) Biodegradation of Commercial Gasoline and Optimization of Production of Biosurfactants. (Poster) 24th Brazilian Congress of Microbiology - Brasília - Brazil

Morales D, Weissheimer MIK, Oliveira NM, Scaramal A, Jacques R, Peralba MCR & Bento FM (2007) Biodegradation of Commercial Gasoline (BTX fraction and ethanol) and Production of Biosurfactants by *Pseudomonas aeruginosa*. (Poster) XXXI Brazilian Congress of Soil Science - Gramado - Brazil.

Teixeira AS, **Morales D**, Peralba MCR, Jacques RJS, Camargo FA de O & Bento FM (2006) Preferential degradation of ethanol in gasoline by soil microorganisms (Poster) Second Brazilian Symposium on Petroleum Biotechnology Old and New Energy Sources - Rio Grande do Norte - Brazil

de Lima-Morales D, Chaves-Moreno D, Jarek M, Vilchez-Vargas R, Jauregui R & Pieper DH (2013) Draft genome sequence of *Pseudomonas veronii* strain 1YdBTEX2. *Genome Announc* **1**: e00258-13.